Regulatory network analyses highlight transcription factor cooperation in haematopoiesis and leukaemogenesis



Joe Ritchie Harman

Hertford College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2022

Acknowledgements

First and foremost, I would like to thank Prof. Tom Milne and Prof. Marella de Bruijn for being fantastic and supportive mentors. Working in Marella's lab gave me my first taste of academia, and helped me develop the critical groundwork for being an academic scientist. Tom has been an incredible guide in navigating my DPhil project, and has always pushed me to be imaginative and driven in my work.

The Milne lab has been home to many lovely colleagues, and I am grateful to work alongside all of them. Nick Crump has taught me many techniques and pushed my critical thinking. I will always remember the late-night lab chats, and spurious analysis sessions. I-Jun Lau has been kind and supportive, and always ready to talk science. To Nick and I-Jun: I'll never stop singing "do you like tdTomato". Alastair Smith has been good to talk to (and complain to) about coding, though I won't stop defending R over Python. Vassi Sharlandjieva, Nicole Jackson and Catherine Chahrour have been bright personalities in the lab, and have brought a little life back into work-life balance. Ross Thorne left the lab before I joined, but his initial work on networks was critical to my thesis, and his viva presentation was immensely inspirational. Outside the lab, Siobhan Rice and Simona Valletta were a fantastic help with mouse experiments.

In the de Bruijn lab, Vincent Frontera was always entertaining to work with, and taught me a lot about flow cytometry. I'm thankful for Christina Rode, who patiently taught me many basic lab techniques, and was always a positive personality. Dominic Owens was a great colleague, who helped me with his CRISPR and capture-C knowledge, and always had a great story to tell. Emanuele Azzoni taught me how to perform whole mount imaging, which are not only scientifically useful, but are artistically beautiful. Lucas Greder performed experiments that have been the seed for my network analyses. Giorgio Anselmi and Alessandro Cavallo have always been ready and willing to talk science, particularly single cell analyses, and have offered great insight. Matt Nicholls has been meticulous in organising the lab, and I am grateful for the times he has kept me company in the mouse facility. Naeema Mehmood has been lovely to talk to - and I appreciate being able to talk about the latest Marvel films!

Most importantly, I would like to thank my family. My mum, Angela, pushed me from a young age to do well academically. I'm sure I can only do maths because we recited the times table so often! My dad, Ritchie, has always supported me, and I inherited my love of fancy tech from him. My sister (the "cooler" sibling), Rachael, has been consistently supportive.

Declaration

I declare that the work presented in this thesis is my own unless otherwise stated, for example where my analyses used data from published work or data from collaborators. This thesis has not been submitted, either partially or in full, for another qualification of this University, or for a qualification at any other institution.

Much of the work in chapter 4 has been published in Harman *et al.*, 2021. I wrote the manuscript, analysed the data, and performed experiments in this publication and I declare that no work performed solely by others has been included, unless otherwise stated.

Collaborators and their contributions to this work:

For computational analyses presented in Chapters 3 and 4, I used experimental data generated by previous/current members of the de Bruijn and Milne labs, alongside published datasets. The sources of experimental data are indicated in the figure legends and in the main body of text. Unpublished data was sourced as follows:

Fig. 3.1 Unpublished mini-bulk RNA-seq and ATAC-seq data on cells undergoing EHT in the mouse embryo (de Bruijn group, L. Greder and G. Swiers) was used where I generated and explored an integrated network model of EHT.

Fig. 3.8 Published tiled capture-C (Owens *et al.*, 2022) was used to complement and extend the integrated EHT network model.

Fig. 3.15 Unpublished Capture-C and qRT-PCR data (de Bruijn group, D. Owens and L. Greder) was used to complement findings from the EHT network.

Fig. 4.1 Analysis performed by R. Thorne was included to introduce the hypothesis for chapter 4.

Fig. 4.2 Unpublished MLL-N ChIP-seq data from patient samples (Milne group, N. Crump) was analysed and included to validate ChIP-seq peak calling.

Fig. 4.10 Unpublished nascent RNA-seq data following *RUNX1* knockdown (Milne lab, M. Tapia) was used to generate a network model of RUNX1 in MLL-AF4 leukaemia.

Fig. 4.8 Analysis performed by T. Wilson on a published CRISPR essentiality screen (DepMap, Avana 21Q1, Doench *et al.*, 2016; Meyers *et al.*, 2017), under my supervision.

Fig. 4.17, 4.20, 4.19, 4.18 Single cell multiome data were generated in a collaboration between the Milne and Roy labs, involving myself, A. Smith, N. Elliott, and S. Rice. Data analysis was done by A. Smith and me: A. Smith contributed the pre-processing and quality control filtering of the ATAC-seq data. The rest of the analysis was performed by me.

Fig. 5.2 Mouse experiments (GMP isolation) were performed with assistance fom S. Valletta and S. Rice. The MLL-AF9 mCherry retroviral construct and $Myb^{\mathrm{MRE/MRE}}$ mouse models used in these experiments was designed and generated by I. Lau (Milne group).

List of publications

*Joint first authors indicated with **

Harman JR*, Thorne R*, Jamilly M, Tapia M, Crump NT, Rice S, Beveridge R, Morrissey E, de Bruijn MFTR, Roberts I, Roy A, Fulga TA, Milne TA. A KMT2A-AFF1 gene regulatory network highlights the role of core transcription factors and reveals the regulatory logic of key downstream target genes. Genome Res. 2021 Jun 4;31(7):1159–73

Harman JR*, Greder L*, Swiers G, Owens DDG, Suciu M, Downes D, Telenius J, Frontera V, Antoniou S, Repapi E, Azzoni E, de Bruijn MFTR. A gene regulatory network of the endothelial-to-hematopoietic transition reveals a Runx1:Ikzf1:Notch feed-forward loop. *Manuscript in preparation.*

Fidyt K, Pastorczak A, Cyran J, Crump NT, Goral A, Madzio J, Muchowicz A, Poprzeczko M, Domka K, Komorowski L, Winiarska M, **Harman JR**, Siudakowska K, Graczyk-Jarzynka A, Patkowska E, Lech-Maranda E, Mlynarski W, Golab J, Milne TA, Firczuk M. Potent, p53-independent induction of NOXA sensitizes MLL-rearranged B-cell acute lymphoblastic leukemia cells to venetoclax. Oncogene. 2022 Mar;41(11):1600-1609

Owens DDG, Anselmi G, Oudelaar AM, Downes DJ, Cavallo A, **Harman JR**, Schwessinger R, Bucakci A, Greder L, de Ornellas S, Jeziorska D, Telenius J, Hughes JR, de Bruijn MFTR. Dynamic Runx1 chromatin boundaries affect gene expression in hematopoietic development. Nat Commun. 2022 Feb 9;13(1):773

Azzoni E, Frontera V, Anselmi G, Rode C, James C, Deltcheva EM, Demian AS, Brown J, Barone C, Patelli A, **Harman JR**, Nicholls M, Conway SJ, Morrissey E, Jacobsen SEW, Sparrow DB, Harris AL, Enver T, de Bruijn MFTR. The onset of circulation triggers a metabolic switch required for endothelial to hematopoietic transition. Cell Rep. 2021 Dec 14;37(11):110103

Godfrey L, Crump NT, O'Byrne S, Lau IJ, Rice S, **Harman JR**, Jackson T, Elliott N, Buck G, Connor C, Thorne R, Knapp DJHF, Heidenreich O, Vyas P, Menendez P, Inglott S, Ancliff P, Geng H, Roberts I, Roy A, Milne TA. H3K79me2/3 controls enhancer-promoter interactions and activation of the pan-cancer stem cell marker PROM1/CD133 in MLL-AF4 leukemia cells. Leukemia. 2021 Jan;35(1):90-106

Owens DDG, Caulder A, Frontera V, **Harman JR**, Allan AJ, Bucakci A, Greder L, Codner GF, Hublitz P, McHugh PJ, Teboul L, de Bruijn MFTR. Microhomologies are prevalent at Cas9-induced larger deletions. Nucleic Acids Res. 2019 Aug 22;47(14):7402-7417

Godfrey L, Crump NT, Thorne R, Lau IJ, Repapi E, Dimou D, Smith AL, **Harman JR**, Telenius JM, Oudelaar A, Downes DJ, Vyas P, Hughes JR, Milne TA. DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. Nature Communications. 10, 2803 (2019).

Conference attendance

Presentation: "Reconstruction of an EHT gene regulatory network identifies a Runx1:Ikzf1:Notch feed-forward loop", International Society for Experimental Hematology. Online, August 2021.

Poster: "RUNX1 gene regulatory network in leukaemia", International RUNX conference. Seoul National University, Seoul, Korea. June 2019.

Contents

List of Figures **xi**

List of Tables **xiv**

List of Abbreviations **xv**

List of terminology definitions **xvii**

Abstract **xix**

1 Introduction **1**

1.1 Mechanisms of gene regulation 2
1.1.1 Chromatin organisation 2
1.1.2 Epigenetic regulation of gene expression 3
1.1.3 Transcription factor structure and function 4
1.2 Gene regulatory network models as tools to investigate gene regulation 7
1.2.1 Boolean networks 8
1.2.2 Gene co-expression networks 8
1.2.3 ChIP-seq and TF binding motif networks 9
1.2.4 Network features 10
1.2.5 Measures of centrality 11
1.2.6 GRN motifs 13
1.2.7 Model systems to apply network analyses 15
1.3 The endothelial-to-haematopoietic transition 16
1.3.1 Transcriptional regulation of EHT 18
1.4 *MLL*-rearranged leukaemia 19
1.4.1 Structure and function of the MLL fusion protein complex . 21
1.4.2 Transcriptional impact of MLL fusion proteins 24
1.5 Thesis overview and aims 26

2 Materials and methods **29**

2.1 Mouse husbandry 30
2.2 Embryo studies 30
2.2.1 Timed matings and embryo collection 30
2.2.2 Embryo tissue dissection and cell sorting 31
2.2.3 Whole mount immunofluorescence imaging 32
2.3 Cell culture 33
2.3.1 Culturing mESCs 33

2.3.2 Culturing THP1, SEM, and RS4;11 cell lines 34
2.3.3 mESC differentiations 34
2.4 MLL-AF9 transduced GMP experiments 35
2.4.1 Bone marrow GMP isolation 35
2.4.2 Transduction of mouse GMPs 36
2.4.3 MethoCult expansion of transduced GMPs 37
2.5 siRNA knockdowns 38
2.6 Colony-forming assays 38
2.7 Drug treatments and dose-response curve analysis 39
2.8 Western blot analysis 39
2.9 Gene expression analysis by qRT-PCR 40
2.9.1 Total RNA extraction and cDNA synthesis 40
2.9.2 qRT-PCR analysis 41
2.10 RNA-seq methods 42
2.10.1 Poly-A RNA-seq library preparation 42
2.10.2 Nascent RNA-seq library preparation 43
2.10.3 Library preparation from embryonic populations 43
2.11 Chromatin immunoprecipitation (ChIP) methods 44
2.11.1 ChIP 44
2.11.2 ChIP qPCR 45
2.11.3 ChIP-seq library preparation 46
2.11.4 ChIPmentation library preparation 47
2.12 ATAC-seq library preparation 48
2.13 RNA+ATAC multiome library preparation 49
2.14 Computational analysis 50
2.14.1 RNA-seq analysis 50
2.14.2 ChIP-seq and ATAC-seq analysis 52
2.14.3 Dimensionality reduction methods 53
2.14.4 Single-cell sequencing analysis 53
2.14.5 Pseudotime analysis 55
2.14.6 Pathway enrichment 56
2.15 GRN construction and analysis 56
2.15.1 TF binding motif analysis 56
2.15.2 EHT GRN methodology 56
2.15.3 MLL-AF4 leukaemia GRN methodology 57
2.15.4 MLL-AF9 GMP leukaemogenesis GRN methodology 58
2.15.5 Centrality calculation 59
2.15.6 TF Co-interaction analysis 60
2.16 Statistics 60
2.17 Sourcing of patient samples 60
2.18 Sourcing of published datasets 61

3 A regulatory network model underlying the endothelial-to-haematopoietic transition 62
3.1 Introduction and aims . . . 63
3.2 Profiling of EHT sub-populations reveals dynamic transcription and chromatin accessibility changes . . . 65
3.2.1 Different EHT expression signatures drive differentiation, morphological, and signalling processes . . . 66
3.2.2 Transition-specific regulation . . . 70
3.3 Generating, and characterising a GRN model of EHT . . . 73
3.3.1 Central regulators and targets of the EHT GRN . . . 75
3.4 *Runx1* regulators and target genes . . . 75
3.4.1 Cis-regulatory landscape of the *Runx1* locus . . . 76
3.4.2 Predicted regulators of *Runx1* expression . . . 78
3.4.3 Predicted targets of Runx1 . . . 81
3.5 Runx1 cooperates with TFs to drive EHT processes . . . 83
3.6 Potential Runx1 driven FFLs in EHT . . . 88
3.6.1 A Runx1 and Ikzf1 driven FFL regulates Notch pathway genes 90
3.7 Maf and Lef1 cooperation may drive cytoskeletal remodelling and EMT processes . . . 94
3.8 Conclusions and discussion . . . 98

4 MLL-AF4 and RUNX1 driven regulatory networks in *MLL*r leukaemia 104
4.1 Introduction and aims . . . 105
4.2 Generating and characterising a GRN model of MLL-AF4 driven leukaemia . . . 108
4.3 Integrating patient expression data highlights regulatory modules, and a set of "core" MLL-FP TFs . . . 113
4.3.1 Regulatory modules of the GRN . . . 115
4.3.2 High centrality nodes represent a core MLL-FP program . . . 116
4.4 Validating the importance of central regulatory hubs . . . 119
4.4.1 Essentiality screens validate the importance of several central TFs . . . 119
4.4.2 Stress is a better indicator of essentiality than degree centrality 121
4.5 MLL-AF4 and RUNX1 cooperate to regulate gene expression . . . 123
4.5.1 Generating, and characterising a RUNX1 driven GRN model of leukaemia . . . 123
4.5.2 RUNX1 and MLL-AF4 GRNs intersect at leukaemogenic processes . . . 125
4.6 MLL-AF4 and RUNX1 direct both FFL and TF cascade regulation 126

4.6.1 Identifying FFL and cascade regulatory circuits 126
4.6.2 Correlation between MLL-AF4 and RUNX1 ChIP-seq binding profiles 130
4.7 TF cooperation at MLL-AF4 targets drives expression and promotes leukaemogenesis 133
4.8 MLL-AF4 and cooperative TFs co-opt and drive normal fetal circuits. 136
4.9 Conclusions and discussion 145

5 A regulatory network driving MLL-AF9 GMP leukaemogenesis 151
5.1 Introduction and aims 152
5.2 Experimental design to characterise MLL-AF9 GMP leukaemogenesis 155
5.3 Gene expression and chromatin accessibility adapts over time, while MLL-AF9 binding is stable. 157
5.4 Generating a GRN model of MLL-AF9 GMP leukaemogenesis . . . 160
5.5 Sub-networks can describe TFs driving expression modules 162
5.6 *Myc* upregulation occurs at week 3, after activation of multiple predicted regulators 164
5.7 Myb and MLL-AF9 are predicted to co-interact with a range of TFs to drive leukaemogenesis 166
5.8 Myb and Runx1 may cooperate to drive expression of key regulatory hubs. 168
5.9 Myb perturbation may cooperate with Runx1 to impact leukaemogenesis and dysregulate essential targets. 170
5.10 Conclusions and discussion 174

6 General discussion 179
6.1 TF cooperation underpinning normal and aberrant cell processes . . 180
6.2 The deregulation of normal GRNs in *MLL*r leukaemia. 182
6.3 Future work 185
6.3.1 Validating predicted interactions 185
6.3.2 Investigating TF cooperativity 186
6.3.3 Improving GRN analyses 186

Appendices

A Appendix 190

References 209

List of Figures

1.1 Illustration of network centrality measures. 12
1.2 Illustration of network motifs. 14
1.3 Schematic of EHT at an E10.5 dorsal aorta. 16
1.4 Incidence of ALL and infant *MLL*r fusion partners. 20
1.5 Schematic of *MLL* translocations, and wild type and MLL fusion protein structure. 22

3.1 Overview of profiled EHT populations. 65
3.2 EHT principal components define switch and transient gene regulation. 67
3.3 Transient regulation activates signalling and morphological changes while switches drive a haematopoietic phenotype. 69
3.4 Analysis of enhancer-promoter pairs suggests stage-specific epigenetic regulation. 71
3.5 *CD109* expression and protein localisation in E8.5 embryos. 72
3.6 Constructing a GRN model of EHT. 74
3.7 Centrality analysis predicts TF drivers of EHT modules. 76
3.8 Chromatin accessibility profiles and predicted enhancers across the *Runx1* locus. 77
3.9 Predicted upstream regulation of *RUNX1*. 80
3.10 Transition specific TF sub-networks highlight predicted and Runx1 targets. 82
3.11 TFs cooperate in clusters to drive different EHT processes. 85
3.12 Gene expression correlation between CoInt-1 TFs in scRNA-seq data. 87
3.13 Motif enrichment of Runx1 co-interacting TFs across ATAC-seq accessibility modules. 89
3.14 Runx1 and Ikzf1 cooperate to repress the Notch pathway through a Runx1:Ikzf1:*Hes1* FFL. 91
3.15 Validation of Runx1:*Ikzf1* interaction. 93
3.16 The transient co-interaction module is predicted to drive cytoskeletal rearrangement through Maf:Lef1 cooperative action. 96
3.17 Summary model of Runx1 and Ikzf1 cooperation to drive *Hes1* repression. 102

4.1 MLL-AF4 regulation cannot be explained by direct binding alone. . 106
4.2 ChIP-seq data is of sufficient quality for GRN analysis. 109
4.3 Constructing an MLL-AF4 centred GRN model of MLL-AF4 leukaemia. 110
4.4 Validation of the MLL-AF4 GRN. 112

4.5 The MLL-AF4 GRN can be split into regulatory modules based on patient gene expression. 114
4.6 Patient expression modules reveal a set of core, central nodes. . . . 117
4.7 The MLL-AF4 GRN predicts core interactions representative of both MLL-AF4 ALL and MLL-AF9 AML. 118
4.8 Essential genes. 120
4.9 Centrality analysis of MLL-AF4 GRN. 122
4.10 Establishing a RUNX1 centred GRN in SEM cells. 124
4.11 The overlap between MLL-AF4 and RUNX1 GRNs is enriched for processes promoting leukaemogenesis. 126
4.12 Network motifs describing interplay between RUNX1 and MLL-AF4. 127
4.13 RUNX1 inhibitor treatment reduces the IC50 concentration of venetoclax. 130
4.14 MLL-AF4 establishes an environment encouraging TF binding and FFL formation. 131
4.15 GRN predicted cooperative regulation of *MYC* and *BCL2*. 134
4.16 RUNX1 and MYB cooperate to drive *MYC* and *BCL2* expression. . 135
4.17 Sc-multiome trajectory analysis of FBM and FL cells recapitulates three fetal lineages. 138
4.18 pre-pro-B/pro-B biased B-lymphopoiesis associated genes are enriched in MLL-AF4 ALL blasts. 139
4.19 The MLL-AF4 GRN drives activation of pre-pro-B/pro-B biased genes, and may repress mature B-cell circuits. 141
4.20 MLL-AF4 and RUNX1 regulate a specific isoform of the fetal B-progenitor TF, *ERG*. 143
4.21 Summary model of how MLL-AF4 binding may alter TF activity. . 148
4.22 Summary model of RUNX1 and MYB cooperation to drive *MYC* and *BCL2* expression. 149
4.23 Summary model describing MLL-AF4 activation of developmental circuits. 150

5.1 Schematic of $Myb^{\mathrm{MRE/MRE}}$ engineered circuit. 152
5.2 Experimental plan for MLL-AF9 transformation of GMPs and MethoCult expansion. 155
5.3 MLL-AF9 GMP replating validation. 156
5.4 RNA and chromatin accessibility adapts from W1 to W2/3, yet MLL-N profiles are stable by W1. 158
5.5 Constructing a GRN model of MLL-AF9 GMP transformation. . . 161
5.6 Generation and visualisation of MLL-AF9 GMP GRN sub-networks. 163

5.7 *Myc* is not fully expressed until week 3, potentially through additional TF regulators. . . . 165
5.8 TF co-interaction drives differentiation, growth, and signalling pathways. . . . 167
5.9 Myb and Runx1 co-interact at multiple loci. . . . 169
5.10 Myb and Myb:Runx1 target genes are more sensitive to *Myb* perturbation by the $Myb^{MRE/MRE}$ circuit. . . . 171
5.11 *Myb* is highly correlated with essential genes impacted by the $Myb^{MRE/MRE}$ circuit. . . . 172
5.12 Summary model of how circuit activity progressively changes in response to MLL-AF9. . . . 175

6.1 Hypothetical model describing Runx1 regulation of the transient EHT network. . . . 181
6.2 Summary model describing possible mechanisms by which the normal GRN is affected in *MLL*r leukaemia. . . . 183

A.1 Relationship between different centrality measures. . . . 191
A.2 Vector map of MSCV MLL-AF9 PGK-mCherry. . . . 192
A.3 Quality control plots for EHT RNA-seq and ATAC-seq. . . . 193
A.4 FastQC and mapping statistics for unpublished datasets in chapter 4. 194
A.5 RNA+ATAC multiome quality control plots. . . . 196
A.6 Sc-multiome lineage markers defining B progenitor populations. . . . 197
A.7 Tapestation traces for MLL-AF9 GMP MethoCult replating. . . . 198
A.8 FastQC and mapping statistics for MLL-AF9 GMP MethoCult replating. . . . 199
A.9 Flow diagram explaining the integration of RNA-seq and ATAC-seq data to generate the EHT GRN model. . . . 200
A.10 Flow diagram explaining the processing of ATAC-seq, RNA-seq and ChIP-mentation data to generate the MLL-AF9 GMP GRN. . . . 201
A.11 Flow cytometry panels for EHT haematopoietic population isolation. 202
A.12 Published ChIP-seq validation of TF motifs. . . . 203
A.13 *Runx1* correlation with predicted Runx1 targets, across EHT GRN modules. . . . 204
A.14 Functional validation of the Runx1:Ikzf1:*Hes1* FFL. . . . 205
A.15 Sc-multiome erythroid and myeloid (neutrophil) lineages. . . . 206
A.16 MLL-AF4 GRN enrichment in normal FBM and FL. . . . 207
A.17 *Myc* enhancer annotations in MLL-AF9 GMPs. . . . 208

List of Tables

2.1 Antibodies used for flow cytometry analysis and FACS sorting of EHT cell populations. 32
2.2 Antibodies used for whole-mount immunofluorescence. 33
2.3 Antibodies used for GMP isolation. 36
2.4 Commercial siRNA used for perturbation experiments. 38
2.5 Antibodies used for western blot analysis 40
2.6 Primers and TaqMan probes used for gene expression qRT-PCR analysis 42
2.7 Antibodies used for ChIP experiments. 45
2.8 Primers used for ChIP qPCR analysis. 46
2.9 Antibody panel for sorting fetal bone marrow and liver donor cells. 50
2.10 Antibody panel for sorting ALL blast samples. 51
2.11 Published datasets used in this thesis, detailing publication and GEO accession numbers. 61

4.1 Patient and healthy donor samples used for single cell RNA+ATAC multiome analysis 137

List of Abbreviations

23GFP . . . . . +23 *Runx1* enhancer GFP reporter

AGM . . . . . Aorta-Gonad-Mesonephros

ALL . . . . . Acute Lymphoblastic Leukaemia

AML . . . . . Acute Myeloid Leukemia

ATAC . . . . . Assay for Transposase-Accessible Chromatin

bZIP . . . . . basic region leucine zipper

chALL . . . . . Childhood ALL

ChIP . . . . . Chromatin ImmunoPrecipitation

CPM . . . . . Counts Per Million

CTG . . . . . CellTiter-Glo (CTG)

DAE . . . . . Differentially Accessible Element

DBD . . . . . DNA Binding Domain

DEG . . . . . Differentially Expressed Gene

DOT1L . . . . . Disruptor Of Telomeric Silencing 1-Like

EC . . . . . Endothelial Cell

EHT . . . . . Endothelial-to-Haematopoietic Transition

E-P . . . . . Enhancer-Promoter association

EMT . . . . . Epithelial-to-Mesenchyme Transition

FBM . . . . . Fetal Bone Marrow

FC . . . . . Fold Change

FCS . . . . . Fetal Calf Serum

FFL . . . . . Feed-Forward Loop

FL . . . . . Fetal Liver

GCN . . . . . Gene Co-expression Network

GMP . . . . . Granulocyte-Monocyte Progenitor

GRN . . . . . Gene Regulatory Network

HE . . . . . Haemogenic Endothelium

HDAC . . . . . Histone DeAcetylase Complex

HPC . . . . . Haematopoietic Progenitor Cell

HSC . . . . . Haematopoietic Stem Cell

HSPC . . . . . Haematopoietic Stem and Progenitor Cell

IAC . . . . . Intra-Aortic Cluster

iALL . . . . . Infant ALL

KD . . . . . Knock-Down

LMPP . . . . . Lympho-Myelo Primed Progenitors

mESC . . . . . mouse Embryonic Stem Cells

MLL . . . . . Mixed-Lineage Leukaemia

***MLL*r** . . . . . *MLL*-rearranged

MLL-FP . . . . . MLL Fusion Protein

MRE . . . . . miRNA Response Element

NT . . . . . Non-Targeting

PAS . . . . . Para-Aortic Splanchnopleura

PCA . . . . . Principal Component Analysis

Pol II . . . . . RNA Polymerase II

PWM . . . . . Position Weight Matrix

sc-multiome . . . Single cell RNA+ATAC multiome

SEC . . . . . Super Elongation Complex (SEC)

TAD . . . . . Topologically associated domain

TF . . . . . Transcription Factor

TPM . . . . . Transcripts Per Million

TSS . . . . . Transcription Start Site

uCpG . . . . . Unmethylated CpG

UMAP . . . . . Uniform Manifold Approximation and Projection

VU . . . . . Vitelline and Umbilical

WNN . . . . . Weighted Nearest Neighbour

wt-MLL . . . . . wild type MLL

List of terminology definitions

Bound gene

A gene is considered bound if a ChIP-seq peak for a TF is found annotated to the nearest expressed promoter.

Co-interaction

Cases where two TFs commonly regulate the same enhancer or promoter regions in a GRN model. This definition does not imply cooperation.

Definitive haematopoiesis

The usage in this thesis refers to the generation of haematopoietic progenitors that persist in adult life, originating from either the yolk sac or the AGM. Historically, the yolk sac has previously been considered the site of primitive haematopoiesis.

Degree centrality

Network statistic. The total number of edges connecting to an individual node (gene/protein).

Edge

Connection within a network. Two nodes are connected by an edge, which can be directional. Represented in this thesis as arrows, where the arrow point indicates the direction, such as TF regulation directed to a gene target.

MLL-N/AF4-C

When referring to MLL-N or AF4-C in ChIP-seq experiments, this refers to antibodies generated recognising the N or C terminus of the protein. MLL-N antibodies will bind to both wt-MLL and MLL-AF4. This project assumes MLL-N binding profiles reflect MLL-AF4 binding, as the majority of the DNA association domains are retained on the N-terminus portion of MLL. Robust MLL-AF4 peak calls were determined by MLL-N and AF4-C overlaps.

Network motif

The usage of the term "network" or "GRN" motif (as opposed to TF binding motif) refers to FFL and cascade motifs. These being specific 3-node network configurations that commonly occur as sub-structures in networks.

Node

Unit of a network. In this thesis this is used to refer to genes and their protein products interchangeably, depending on whether they are regulating (TF proteins) or targets (genes).

Stress and betweenness centrality

Similar network statistics that quantify how often a node is on a shortest path between other nodes of the network. Shortest paths are calculated between all possible pairs of nodes. Stress and betweenness are measures of how frequently an individual node is part of a shortest path. Stress is a quantification of shortest paths, while betweenness is represented as the proportion of all shortest paths in the network that pass through the node in question. As such, these measures can be thought to measure "bottle-necks" between two network regions.

TF binding motif

The usage of the term TF (DNA) binding motif (as opposed to network motif) refers to DNA sequences that particular TFs have an affinity to bind to.

Abstract

The regulation of gene expression is a complex process that confers cell identity and drives cell differentiation. Transcription factors (TFs) are an important modulator of gene expression and are critical to understanding the regulatory code characterising cell behaviour. However, TFs function in cooperation with one another, acting as a complex combinatorial code to enact gene regulation. Gene regulatory network (GRN) models are used to describe the regulation of genes and to account for this combinatorial code. When the regulatory code underpinning normal cell behaviour goes awry this can lead to abnormal GRNs and the development of disease. This poses two questions: how does TF cooperation drive cellular processes, and how are normal GRNs deregulated in disease?

Haematopoietic stem cells are first established in the developing embryo through a process known as the endothelial-to-haematopoietic transition (EHT). EHT involves dynamic gene regulation across well-defined populations, and is critically dependent on the master TF regulator, RUNX1. EHT is therefore a fascinating system to study TF regulation in driving a complex developmental process. When haematopoietic GRNs are misregulated this can lead to disease. A common childhood leukaemia involves chromosomal rearrangements resulting in novel Mixed Lineage Leukaemia (MLL) fusion proteins that aberrantly drive transcription. MLL-rearranged leukaemias are considered a wholly transcriptional disease, and as such are uniquely suited for studying through GRN approaches.

In this thesis, the study of GRNs underpinning normal and malignant haematopoietic processes revealed insights into the role of TF cooperation. Notably, Runx1 cooperates with TFs to drive repression of a transiently activated network during EHT. In MLL-AF4 leukaemia, MLL-AF4 drives TF cooperation by altering the chromatin environment and encouraging local TF binding, and within this MLL-AF4 context RUNX1 activity is biased towards activation. RUNX1 was also found to cooperate with MYB to synergistically activate *BCL2* and *MYC*, promoting leukaemogenesis.

1

Introduction

Contents

1.1 Mechanisms of gene regulation 2
1.1.1 Chromatin organisation 2
1.1.2 Epigenetic regulation of gene expression 3
1.1.3 Transcription factor structure and function 4
1.2 Gene regulatory network models as tools to investigate gene regulation 7
1.2.1 Boolean networks 8
1.2.2 Gene co-expression networks 8
1.2.3 ChIP-seq and TF binding motif networks 9
1.2.4 Network features 10
1.2.5 Measures of centrality 11
1.2.6 GRN motifs 13
1.2.7 Model systems to apply network analyses 15
1.3 The endothelial-to-haematopoietic transition 16
1.3.1 Transcriptional regulation of EHT 18
1.4 *MLL*-rearranged leukaemia 19
1.4.1 Structure and function of the MLL fusion protein complex 21
1.4.2 Transcriptional impact of MLL fusion proteins 24
1.5 Thesis overview and aims 26

Regulation of gene transcription is a core aspect of cell behaviour. While all cells contain the same set of DNA, it is the regulation of genes that code for different cell identities and orchestrate developmental processes. This regulatory code can sometimes run awry, leading to the development of disease (Baxter *et al.*, 2014; Lee *et al.*, 2013; Thoms *et al.*, 2019). The regulatory code that drives specific cell processes, and how these are disrupted in disease, are areas of active study. Gene regulation is dependent on cellular context, and this complexity offers many questions as to how specific processes are driven or are misregulated in disease. It is possible for complex gene regulation to be decoded through the use of network models (Levine *et al.*, 2005), and the application of such networks may lead to insights into the regulation of specific processes.

1.1 Mechanisms of gene regulation

1.1.1 Chromatin organisation

Gene transcription is a complex process, that is not yet fully understood. Factors that regulate gene expression are dependent on the structure of chromatin, which is therefore the first level of transcriptional regulation. Genomic DNA is tightly packaged into chromatin, of which the basic unit is the nucleosome. A nucleosome consists of ∼145-147 bp of DNA wrapped around a histone octamer core (Luger *et al.*, 1997), which can block regulatory factors and transcriptional machinery from accessing DNA. Each core histone octamer (H2A, H2B, H3, H4) contains an unstructured tail which can be post-translationally modified to regulate DNA accessibility and recruitment of cofactors (Luger *et al.*, 1998; Quina *et al.*, 2006). Studies into the higher order structure of chromatin have identified long-range physical interactions between regions of chromatin. These interactions assemble chromatin into separate structures, known as topologically associated domains (TADs), that categorise DNA into active and inactive regions (Dixon *et al.*, 2012; Lieberman-Aiden *et al.*, 2009).

Within TADs are several genomic features that control gene expression, such as promoters, enhancers, silencers, and boundary elements. Promoters are the sites of transcription initiation, where ribonucleic acid (RNA) polymerase II (Pol II) transcribes messenger RNA (Roeder *et al.*, 1969), starting from a region within a promoter referred to as the transcription start site (TSS) (as reviewed in Cramer, 2019 and Haberle *et al.*, 2018). Promoters often have low basal activity, but are readily modulated by local chromatin modifications or through distal cis-regulatory elements such as enhancers or silencers. Enhancers and silencers are regulatory elements that come into close proximity with promoters within TADs to modulate transcription (Mifsud *et al.*, 2015; Sanyal *et al.*, 2012). Enhancers act to increase the activity of promoters (Banerji *et al.*, 1981), while silencers reduce activity (Brand *et al.*, 1985), as reviewed in Haberle *et al.*, 2018 and Segert *et al.*, 2021. Boundary elements are CTCF enriched sites that demarcate TADs (Braccioli *et al.*, 2019; Dixon *et al.*, 2012), and thus constrain enhancer and promoter interactions between TADs.

1.1.2 Epigenetic regulation of gene expression

Promoters and regulatory elements frequently display various epigenetic modifications, some of which are thought to be functional. Histone tails can be post-translationally modified through the acetylation or methylation of specific amino acid residues (Luger *et al.*, 1998, reviewed in Bannister *et al.*, 2011). Regulatory elements are marked with specific histone modifications that correlate with different functions. For example, active promoters and enhancers are both marked by H3K27ac (histone H3 lysine 27 acetylation), while promoter activity is correlated with H3K4me3 (histone H3 lysine 4 tri-methylation) and enhancers with H3K4me1 (histone H3 lysine 4 mono-methylation) (Barski *et al.*, 2007; Creyghton *et al.*, 2010; Heintzman *et al.*, 2009; Heintzman *et al.*, 2007). Silencers are instead proposed to be marked by H3K27me3 (Huang *et al.*, 2019), though it is increasingly recognised that silencer elements can be bifunctional, and what can be a silencer in one context may act as an enhancer in another (Brand *et al.*, 1987). Histone modifications also lead to the assembly of various cofactors. For example, H3K27ac recruits

bromodomain containing proteins, such as BRD4, which leads to the association of transcriptional machinery that drives gene expression (Dhalluin *et al.*, 1999; Jang *et al.*, 2005; Yang *et al.*, 2005). Another feature of these elements is that the DNA is accessible, as quantified by assays measuring chromatin accessibility (e.g., ATAC-seq), as a result of the chromatin environment (Baniahmad *et al.*, 1987; Parslow *et al.*, 1983; Theisen *et al.*, 1986, reviewed in Gross *et al.*, 1988 and Klemm *et al.*, 2019). DNA can also be methylated, which is typically associated with gene silencing at promoters, though it is debated whether this methylation is a driver of repression, or a consequence of repression (Hashimshony *et al.*, 2003; Lock *et al.*, 1987; Venolia *et al.*, 1983, reviewed in Jones, 2012).

1.1.3 Transcription factor structure and function

Transcription factors (TFs) are critical proteins that modulate gene expression and make up the cell type specific code characterising cell identity and processes, as reviewed in Latchman, 1997. TF proteins are highly modular, usually containing a DNA binding domain (DBD), and an activation domain or repression domain (Brent *et al.*, 1985; Freedman *et al.*, 1988; Hope *et al.*, 1986; Kadonaga *et al.*, 1988; Licht *et al.*, 1990; Lillie *et al.*, 1986; Ma *et al.*, 1987; Sturm *et al.*, 1988, reviewed in Hanna-Rose *et al.*, 1996; Latchman, 1997; Mitchell *et al.*, 1989). The DBD recognises and binds to specific DNA sequences and TFs are generally classified by their DBD structure, such as the basic region leucine zipper (bZIP) family, as discussed in Luscombe *et al.*, 2000 and Weirauch *et al.*, 2011. The use of Chromatin Immunoprecipitation (ChIP) sequencing has allowed the detailed study of DNA sequences underlying TFs, and has been a useful tool in characterising TF binding motifs (Boeva, 2016; Machanick *et al.*, 2011; Mundade *et al.*, 2014). TF binding motifs are characterised as position weight matrices (PWMs), that describe the probability of TF binding to different sequence compositions (Stormo, 2000). As DNA recognition sequences are typically small, there is an abundance of TF binding motifs throughout the genome, and as such the use of TF motifs often overestimates DNA binding. The binding of TFs to their DNA motifs is affected by many factors

and can depend on the specific chromatin environment. An important consideration is that TF binding is also affected by chromatin accessibility, particularly for TFs with low binding affinity (Spitz *et al.*, 2012).

There are several mechanisms by which TFs are thought to activate or repress transcription, primarily centred around their ability to recruit other proteins. The mediator complex is one such factor, that is thought to be generally associated with TF binding and to subsequently regulate Pol II activity (Baek *et al.*, 2006; Johnson *et al.*, 2002, reviewed in Allen *et al.*, 2015). TFs are also known to directly recruit other co-factors. For example, among TFs expressed in haematopoietic cells, IKZF1 may recruit histone deacetylase complexes (HDACs), thus mediating epigenetic repression (Kim *et al.*, 1999a; Marke *et al.*, 2018). MYB is known to recruit CBP and P300, resulting in transcriptional activation through acetylation (Dai *et al.*, 1996; Sandberg *et al.*, 2005). RUNX1 is known to enact both activation and repression, and has been reported to recruit both co-activators (such as p300), as well as co-repressors (TLE1) (Canon *et al.*, 2003; Kitabayashi *et al.*, 1998; Perry *et al.*, 2002; Yamagata *et al.*, 2005). It may be possible for TFs in specific contexts to regulate genes indirectly, as for example, it has been suggested that IKZF1 is capable of sequestering the nucleosome remodelling and histone deacetylase (NuRD) complex (Georgopoulos, 2017). Conversely, it has also been found that recruited co-factors with no intrinsic DBD can alter TF DNA binding preferences, as seen with the TF Cbf1 and cofactor Met28 (Siggers *et al.*, 2011), as well as CBFb and RUNX1 (Ogawa *et al.*, 1993; Wang *et al.*, 1996b).

An important aspect of TF regulation is that they do not act in isolation. In fact, enhancer elements contain TF DNA motifs allowing the binding of a multitude of TF proteins, and the specific combination of regulatory factors gives rise to precise regulation (Bhattacharjee *et al.*, 2013; Halfon *et al.*, 2000; Reiter *et al.*, 2017; Spitz *et al.*, 2012; Yuh *et al.*, 1994). It has also been shown that the expression of specific combinations of TFs can even reprogram cell identity (Batta *et al.*, 2014; Riddell *et al.*, 2014; Schütte *et al.*, 2016; Takahashi *et al.*, 2006; Xie *et al.*, 2004).

The binding affinity of TFs to their motifs are highly context dependent, and this is partly driven by the presence of other TFs, which function together to stabilise DNA interactions (Morgunova *et al.*, 2017; Spitz *et al.*, 2012; Yáñez-Cuna *et al.*, 2012). This effect can be dramatic, as seen where expression of *RUNX1* leads to genome wide redistribution of FLI1 binding (Lichtinger *et al.*, 2012). This TF cooperation can be achieved generally through weak interactions between TFs, but often involves strong dimerization of proteins. This is well known to occur in bZIP TFs, that either form homodimers or heterodimers, often resulting in binding to composite motifs (Miller, 2009; Rodríguez-Martínez *et al.*, 2017). Additionally, RUNX1 naturally autoinhibits its own DNA binding, but interaction with ETS1 blocks this autoinhibition, resulting in increased DNA binding (Goetz *et al.*, 2000; Kim *et al.*, 1999b). Consequently, RUNX1 and ETS1 bind a composite motif, that is known to regulate several important elements, such as antigen receptor enhancers in B-cells (Erman *et al.*, 1998), or T-cell receptor enhancers (Giese *et al.*, 1995; Mayall *et al.*, 1997). Additionally, alongside synergistic cooperation, TFs act at the same sites to confer redundancy, and thus robustness to perturbation. For example, this is seen in T-cell lineage commitment, where Runx1 and Runx3 bind similar sites and act to additively regulate gene expression and have high functional redundancy (Shin *et al.*, 2021). Some TFs lack any regulatory potential on their own, as with the small MAF family (MAFF, MAFG, MAFK). Small MAFs lack an activation domain but form heterodimers with several TFs, and as such direct binding to MAF DNA motifs (Katsuoka *et al.*, 2016). The activity of TFs can also be altered through interactions with other proteins. This is seen with RBPJ, which acts as a transcriptional repressor, but forms a transactivating complex with the NOTCH intracellular domain (NICD) (Kojika *et al.*, 2001).

It is clear that the TF driven combinatorial code of gene regulation is critical for understanding biological processes. However, cooperative TF function is complex, and difficult to dissect. Better approaches are required to address this problem,

and the use of networks as a method to synthesise complex data could be used to enhance our understanding of combinatorial TF behaviour.

1.2 Gene regulatory network models as tools to investigate gene regulation

One use of gene regulatory network (GRN) models is to investigate the complex combinatorial code of TF driven gene regulation. GRNs have been used to describe complex biological processes, a field that has been pioneered by work from Eric Davidson starting with the transcriptional network underlying sea urchin development (Davidson *et al.*, 2002; Oliveri *et al.*, 2004; Yuh *et al.*, 1998) and early theory dating back to 1969 (Britten *et al.*, 1969), as reviewed in Rothenberg *et al.*, 2021. GRNs have been applied to describe various systems, such as haematopoietic specification (Goode *et al.*, 2016), T-cell development (Ciofani *et al.*, 2012; Georgescu *et al.*, 2008; Kueh *et al.*, 2012), B-cell development (Boller *et al.*, 2014; Sciammas *et al.*, 2011), neural crest development (Meulemans *et al.*, 2004; Simões-Costa *et al.*, 2015; Williams *et al.*, 2019), and cancers, including lymphoma (de Matos Simoes *et al.*, 2013) and leukemia (Assi *et al.*, 2019). GRNs can be considered a framework for assembling complex data, that can be represented and analysed using network-based methodologies. They can be used to build hypotheses of gene regulation, and acquire an understanding of the association of specific regulators with distinct regulatory modules. It is also a framework that can allow the investigation of specific circuit types, and to explore how different TFs may cooperate at genomic loci (Mangan *et al.*, 2003a; Milo *et al.*, 2002; Shen-Orr *et al.*, 2002). However, it is important to acknowledge that these network models are not definitive, and that they are an alternative approach of viewing existing data. Observations made using GRN models are predictive, and must be experimentally validated.

GRNs in this thesis are used to describe the action of TFs to regulate gene expression. There are many methods to construct such a network model, depending

on the available data, as have been discussed across multiple reviews (Emmert-Streib *et al.*, 2014; Hecker *et al.*, 2009; Karlebach *et al.*, 2008; Mercatelli *et al.*, 2020). As a brief note on terminology used in this thesis, networks (also known as graphs) consist of multiple units referred to as nodes. Connections between these nodes are referred to as edges. In terms of GRNs, nodes are commonly a representation of a gene, and edges describe predicted regulation by TFs. These gene nodes represent both the regulation of the gene (A>B: TF A regulates gene B), but can also represent the protein product of the gene (B>C: protein product B regulates gene C).

1.2.1 Boolean networks

GRNs can be represented using boolean logic (Thomas, 1973), which describes TF activity as discrete on or off states. For example, a boolean computational model has been applied to a sea urchin development GRN in Peter *et al.*, 2012. A variant of this approach has been developed to integrate uncertainty as a probabilistic boolean network, where specific nodes have a probability of performing a specific action (Shmulevich *et al.*, 2002). These approaches, while reductive, facilitate modelling of genetic circuits and prediction of regulatory outputs. However, many TFs show dose-dependent regulation of gene expression (Lie-A-Ling *et al.*, 2018; Mak *et al.*, 2011; Scripture-Adams *et al.*, 2014), and as such boolean methods are not suited for all circuits. Therefore, methods that incorporate continuous gene expression data may be more suitable, such the use of gene expression correlation to infer regulatory interactions.

1.2.2 Gene co-expression networks

Gene expression profiling methods (e.g., RNA-seq) provide detailed information on levels of transcription genome wide, and methods have been developed to incorporate this information into a GRN model. A widely used approach is to create a gene co-expression network (GCN), which relies on calculating correlation between gene expression profiles (Serin *et al.*, 2016). Gene pairs exceeding a correlation

threshold are inferred to interact, or to at least exist within the same network "module". One such implementation is the weighted gene correlation network analysis (Langfelder *et al.*, 2008). As GCNs rely on correlation, this method also predicts positive and negative relationships between genes. However, while correlation methods successfully extract genes that associate with each other, and can infer positive and negative relationships, they cannot imply causation.

Networks can generally be designed to model a steady-state system (static), or a time course following treatment or differentiation (dynamic) (Hecker *et al.*, 2009). Dynamic models inherently exhibit complex gene expression patterns that lend themselves to correlation analyses. As static models may not exhibit significant changes in gene expression, robust correlation scores can be difficult to acquire. It is considered by some that to fully understand a network, components must be systematically perturbed (Hecker *et al.*, 2009; Ideker *et al.*, 2000; Tegnér *et al.*, 2007). Such methods have been used to interrogate GRNs, such as in T-cell development (Zhou *et al.*, 2022), and cancer models (Morgan *et al.*, 2020). Using CRISPR and RNA interference to disrupt specific factors is a powerful tool for investigating the role and impact of TFs within transcriptional networks (Ideker *et al.*, 2000; Tegnér *et al.*, 2007), and can also introduce dynamism into steady state systems.

1.2.3 ChIP-seq and TF binding motif networks

Chromatin accessibility profiling (e.g., DNase-seq, ATAC-seq) and ChIP-seq have also been significant tools in the inference of GRNs. As ChIP-seq provides direct evidence for DNA binding of TFs genome wide, it is a powerful method to predict targets of specific TFs. Several studies have assembled ChIP-seq data into regulatory networks (Schütte *et al.*, 2016; Wade, 2015), which provides a robust collection of interactions. However, ChIP-seq approaches are limited both by low throughput, as only one TF can be probed at a time, and a requirement for high cell numbers. The latter complication has recently been alleviated with tagmentation based techniques, such as ChIPmentation (Crump *et al.*, 2022; Gustafsson *et al.*,

2019). For a general model of all TFs, DNA binding motif-based approaches can be more successful. As TFs bind DNA at regions of accessible chromatin, ATAC-seq can be used to highlight potential regulatory sites. Further, combining ATAC-seq with ChIP-seq for histone modifications, allows the categorisation of sites into enhancers and promoters (Barski *et al.*, 2007; Creyghton *et al.*, 2010; Heintzman *et al.*, 2009; Heintzman *et al.*, 2007). Motif PWMs can be used to determine the probability that these sites harbour specific TF DNA motifs, and as such infer TF regulation (Stormo, 2000). However, as discussed in section 1.1.3, the actual binding of TFs to DNA is complex. Therefore, DNA motif inferred regulation should be reinforced with additional evidence. Further, both ChIP-seq and motif-based approaches do not suggest regulation, but instead indicate that a protein may be bound. Combining this data with GCN methods can establish a more robust model of regulation, with gene expression correlations evidencing positive and negative regulation and association, and ChIP-seq or motifs evidencing direct interaction.

The above approaches have been established using bulk-cell populations. Single-cell based approaches have, in the past, performed poorly for constructing networks (Chen *et al.*, 2018). However, with the development of tailored algorithms (Aibar *et al.*, 2017; González-Blas *et al.*, 2022; Van de Sande *et al.*, 2020) and the advent of RNA-seq and ATAC-seq single cell multiome technology (10X Genomics), single-cell based GRNs may improve.

1.2.4 Network features

There are additional features that aid the interpretation of GRNs. Arguably the most important is the inference of directionality. For example, in a real transcriptional network a TF (TFa) may regulate a TF gene, but the gene product (TFb) might not regulate TFa. Hence directionality will model that TFa regulates TFb, but not vice versa. GCNs are undirected, as causation is not implied. ChIP-seq data and TF DNA motifs can be used to infer directionality, as they suggest direct TF binding. Another parameter is the sign of interactions, where an edge suggests

activating (+) or repressive (-) regulation. This can be inferred from GCNs (positive and negative correlation), or from gene perturbation studies. Following on from this, many networks have been described with the integration of logic gates, as reviewed in Materna *et al.*, 2007. For example, the activation of certain genes may require two TFs (TFa and TFb) to enact positive regulation, before a target gene is expressed (AND logic). Conversely, it may be that the gene can be activated by TFa or TFb (OR logic). This logic can be integrated at specific loci to assemble a detailed understanding of the mechanisms of gene regulation.

1.2.5 Measures of centrality

Network models are inherently complex. A GRN on its own is a format to organise observations and predictions based on genomic data. This does not simplify data, but instead provides a framework by which predictions and calculations can be made. Several calculations have been developed to probe networks and interpret how individual nodes are connected (reviewed in Landherr *et al.*, 2010). The simplest and first established measure is degree centrality, which is the number of connections to or from a node (Fig. 1.1A, Bavelas, 1950; Nieminen, 1974). Higher degree centrality implies greater communication or influence in the network.

More advanced measures of centrality rely on the concept of "shortest paths". This is determined as the smallest possible number of nodes on a path between two other nodes. Closeness centrality is a measure that calculates the average number of shortest paths from the node of interest (NOI) to other nodes in the network (Bavelas, 1950). This assumes that the most important nodes of a network must be generally close to all other nodes in a network. Measures related to this instead quantify how frequently the NOI is an intermediate of a shortest path between all other node pairs in a network. This measure can be represented as the total number of shortest paths (stress centrality), or the fraction of total shortest paths (betweenness centrality), as described in Freeman, 1977 (Fig. 1.1A). By this method, high stress or betweenness centrality nodes are assumed to exert

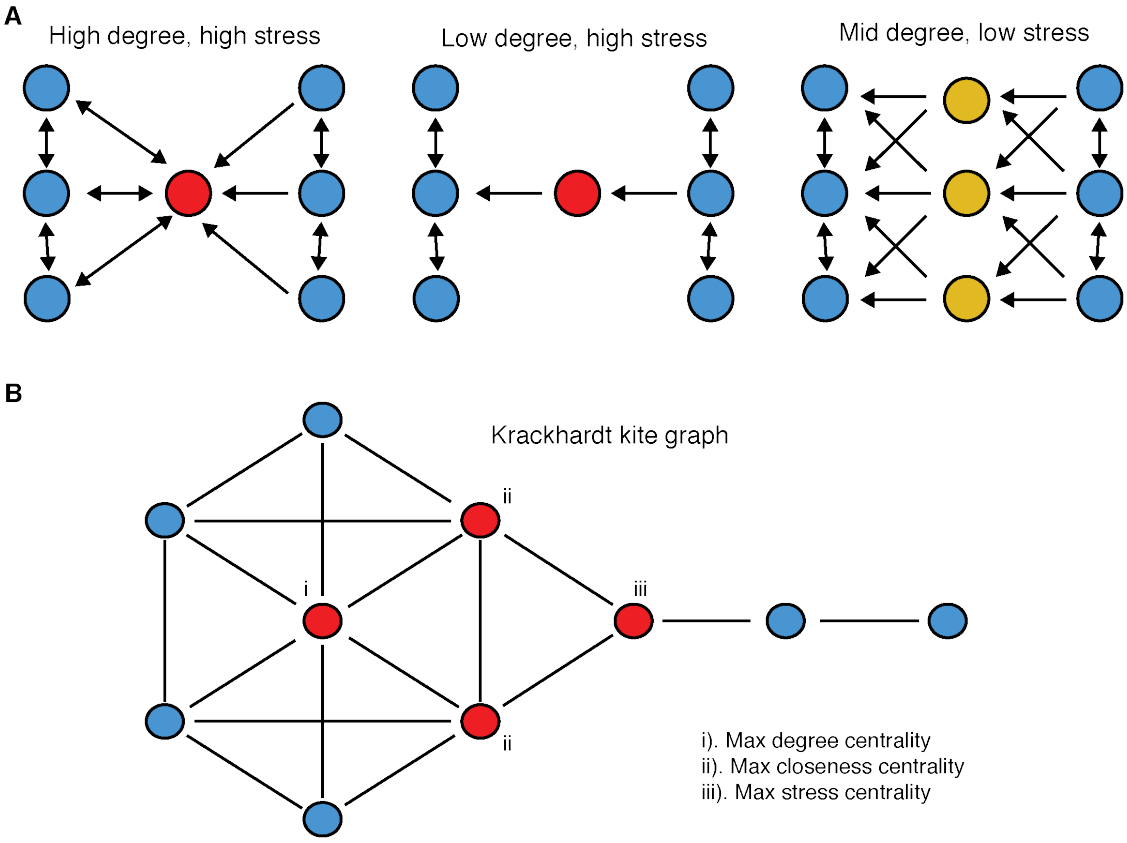


Figure 1.1: Illustration of network centrality measures. (A) Schematic of small example networks, exemplifying different levels of degree and stress centrality. Blue nodes are connected nodes in the networks, red and orange nodes are the measured NOIs. Red nodes display high centrality. Orange nodes display low/mid centrality. **(B)** Krackhardt kite graph (Krackhardt, 1990) illustrating how centrality measures emphasise different network structures.

greater control of communication between other nodes in the network. Different measures of centrality can emphasise different hubs of a network, as exemplified by the Krackhardt kite graph (Fig. 1.1B, Krackhardt, 1990). Therefore, the choice of centrality measures can offer different conclusions. See Appendix A.1 for the relationship between centrality measures.

The relevance of centrality measures in biological networks has been well-studied. For example, studies in *S. cerevisiae*, *C. elegans*, and *D. melanogaster* have shown that high centrality can be associated with essential proteins (Hahn *et al.*, 2005; Jeong *et al.*, 2001; Koschützki *et al.*, 2008). Hahn *et al.*, 2005 also noted that highly central proteins evolved more slowly. Interestingly, it has been suggested that nodes

with low degree centrality but high stress centrality are more likely to represent nodes connecting distinct network modules (see Fig. 1.1A middle panel, Joy *et al.*, 2005; Koschützki *et al.*, 2008). As such, combining centrality measures can reveal useful information on the structure of regulatory networks.

It is important to note that the networks these studies are based on are comparatively simple relative to mouse or human transcriptional networks, and so may not reflect essentiality as well in more complex organisms. Typically, while centrality is often a good indicator of network structure, it is not always a good indicator of essentiality. This is particularly true for TF models, as TF regulatory logic is complex with highly context dependent binding and function as well as regulatory redundancy (see section 1.1.3), that is difficult to capture in the GRN model itself. Nevertheless, investigating the centrality of TF hubs offers interesting insights into the complex structure of TF based networks.

1.2.6 GRN motifs

GRN models are a complex collection of data and observations. While centrality methods highlight specific regulatory hubs, they do not necessarily describe network structure. The Alon lab pioneered an approach to analysing GRNs by breaking graphs down into simple, repeating units (Mangan *et al.*, 2003a; Milo *et al.*, 2002; Shen-Orr *et al.*, 2002). This analysis focused on the identification of 3-4 node patterns, termed network motifs (or GRN motifs), that were enriched in biological networks. By describing networks in such a way, prominent GRN motifs can be probed whilst still preserving the small-scale structure of the network. Of particular note, these studies found one of the most prominently enriched GRN motifs in biological networks to be the feed-forward loop (FFL, Fig. 1.2A, Mangan *et al.*, 2003a; Milo *et al.*, 2002). Another recurring pattern, not specifically enriched in biological networks but important nonetheless, is the TF cascade (Fig. 1.2B), in which a chain of TFs propagate signal to downstream targets (Lee *et al.*, 2002; Rosenfeld *et al.*, 2003).

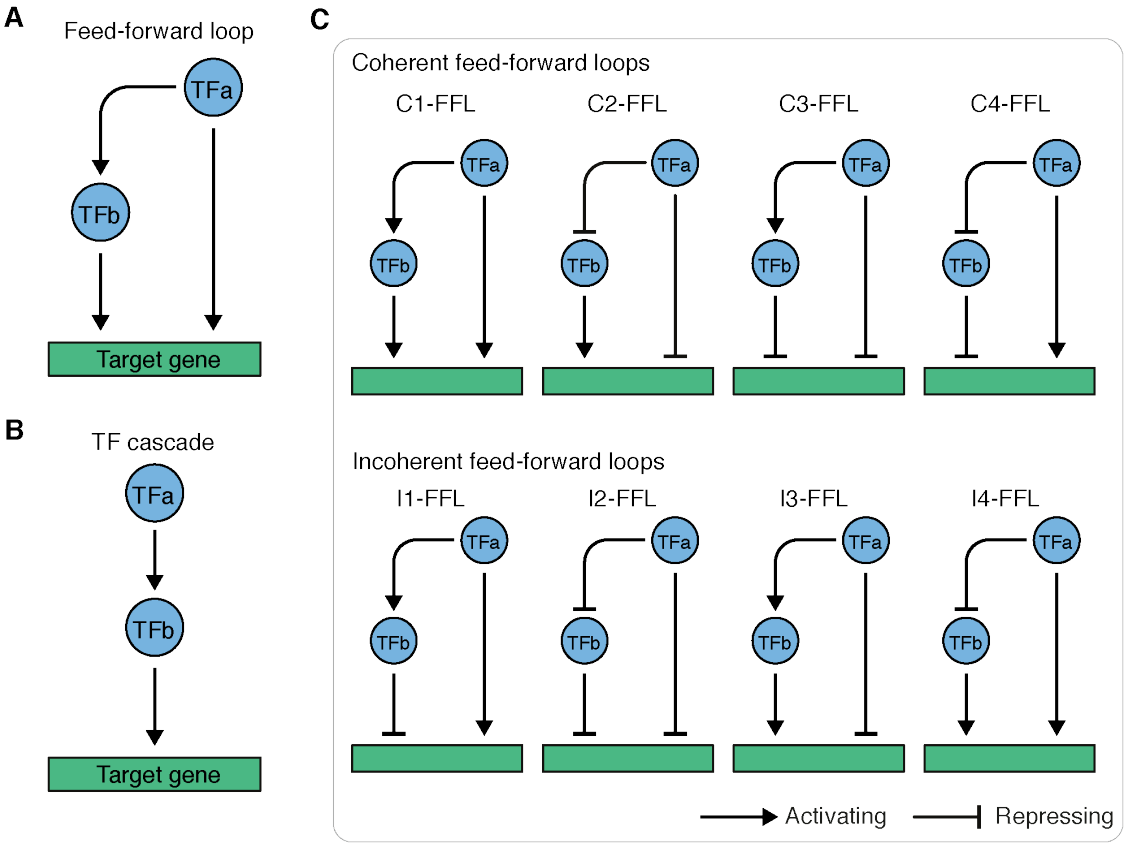


Figure 1.2: Illustration of network motifs. (A) Schematic of a FFL circuit. **(B)** Schematic of a TF cascade. **(C)** Schematic of coherent (top) and incoherent (bottom) FFL sub-types, determined by sign of effect for each interaction. Blue nodes are TF regulators, green boxes are gene targets

Many publications have followed these analyses, focusing on the consequence of network structure on interaction dynamics (Goentoro *et al.*, 2009; Joanito *et al.*, 2018; Mangan *et al.*, 2003a; Mangan *et al.*, 2006). The FFL can be divided into sub-types based on the sign of effect enacted by the TF regulators (Fig. 1.2C, Mangan *et al.*, 2003a). This can be broadly grouped into two sub-types, where the regulatory logic of both TF regulators are the same (coherent) or opposite (incoherent). The C1-FFL and I1-FFL motifs (Fig. 1.2C) were found to be particularly prevalent (Joanito *et al.*, 2018; Mangan *et al.*, 2006). These sub-types can be sub-divided further, as FFLs can be thought to function with AND/OR logic gates. Coherent FFL motifs that exhibit AND logic have been shown to be sign-sensitive delay elements, in that signal processing results in delayed activation going from an OFF to ON state, but rapid deactivation going from ON to OFF (Mangan *et al.*, 2003b).

Coherent FFL motifs are also sign-sensitive, with a delay moving from an ON to OFF state (Kalir *et al.*, 2005). The incoherent I1-FFL has been described as a pulse generator, where the repressive TF (TFb) has a strong effect on gene expression (Basu *et al.*, 2004; Mangan *et al.*, 2003a). This is described as TFa upregulating both TFb and the gene target, followed shortly after by TFb repression of the gene target.

While GRN motifs are a useful tool for probing networks, and can offer insight into gene regulatory dynamics in small-scale circuits, it is important to acknowledge that these motifs do not always translate to biological function (Ingram *et al.*, 2006). Like with measures of centrality, motif structure does not necessarily correlate with importance of a circuit, and to determine regulatory dynamics from GRN motifs detailed biochemical data is needed to support the model, which cannot reasonably be performed genome wide. Despite these drawbacks, GRN motifs are yet another method by which GRN models can be probed, and potentially elucidate key mechanisms of gene regulation.

1.2.7 Model systems to apply network analyses

In this thesis, I have investigated the use of GRN models for investigating the complex behaviour of TF regulation. The application of GRN models to haematopoiesis is an active and evolving field of research (Rothenberg *et al.*, 2021). The diverse range of cell types and lineages, and complex transcriptional regulation involved in haematopoiesis makes this an ideal field to probe the use of TF GRNs. In particular, I aim to use GRNs to describe a process underlying the developmental origin of haematopoietic stem cells (HSCs) known as endothelial-to-haematopoietic transition (EHT) (Bertrand *et al.*, 2010; Boisset *et al.*, 2010; Kissa *et al.*, 2010) (see section 1.3). This is a normal haematopoietic process, yet the disruption of normal circuits can sometimes drive disease. Therefore, I also aim to use GRNs to describe *MLL*-rearranged leukaemias (see section 1.4), to investigate how normal transcriptional networks may be misregulated in disease. Importantly, EHT is dependent on the TF RUNX1 (Cai *et al.*, 2000; Chen *et al.*, 2009; North *et al.*,

1999; Okuda *et al.*, 1996; Wang *et al.*, 1996a), while *MLL*-rearranged leukaemias can involve the overexpression of key TFs, such as *RUNX1* and *MYB*. Therefore, throughout this thesis I aim to anchor GRN analyses on the behaviour of two key TFs: RUNX1 and MYB.

1.3 The endothelial-to-haematopoietic transition

The establishment of the haematopoietic system occurs in utero, and is a complex process that has been extensively explored in mice. In the embryo, haematopoiesis occurs broadly in three waves. The first, primitive, wave involves the generation of primitive erythroid progenitors originating at yolk sac blood islands from embryonic day E7.0 (Palis *et al.*, 1999; Palis *et al.*, 2001). The second wave occurs from E8.25, generating erythro-myeloid progenitors, among others, in the yolk sac (Frame *et al.*, 2013; Palis *et al.*, 1999). The final wave is established in E10.5 embryos, and occurs within the aorta-gonad-mesonephros (AGM) region at the embryonic dorsal aorta, generating the first adult-type HSCs (de Bruijn *et al.*, 2000; Medvinsky *et al.*, 1996; Müller *et al.*, 1994). The second and final waves establish definitive progenitors (originating from the yolk sac as well as the AGM) that persist into adult life (Gentek *et al.*, 2018; Gomez Perdiguero *et al.*, 2015; Yoshimoto *et al.*, 2011; Yoshimoto *et al.*, 2012).

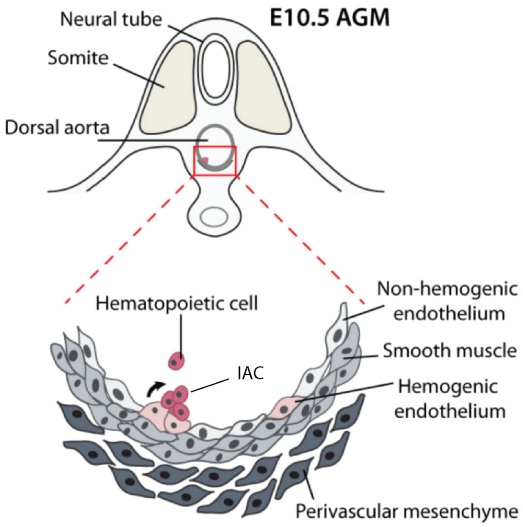


Figure 1.3: Schematic of EHT at an E10.5 dorsal aorta. Schematic of an E10.5 AGM, illustrating the EHT process. Boxed section magnified to highlight cell types. HE is coloured as pink, and haematopoietic cells and IACs as red. Non-HE endothelium is coloured as light grey, smooth muscles as mid grey, and mesenchymal cells as dark grey. *Figure has been taken from Swiers* et al.*, 2013b.*

The initial establishment of haematopoietic stem and progenitor cells (HSPCs) occurs by a process referred to as endothelial-to-haematopoietic transition (EHT) (Bertrand *et al.*, 2010; Boisset *et al.*, 2010; Kissa *et al.*, 2010). This process is best described in the embryonic dorsal aorta, but also occurs in the vitelline and umbilical (VU) arteries and the yolk sac (McGrath *et al.*, 2015; North *et al.*, 1999; Padrón-Barthe *et al.*, 2014; Samokhvalov *et al.*, 2007; Zovein *et al.*, 2008). It starts with the specification of a specialised subset of endothelial cells (EC) known as haemogenic endothelium (HE). Flat HE cells undergo transcriptional and morphological changes, and bud into the lumen of the dorsal aorta to form intra-aortic clusters (IACs) of cells (Fig. 1.3) (Bertrand *et al.*, 2010; Boisset *et al.*, 2010; Kissa *et al.*, 2010; Swiers *et al.*, 2013a).

Runx1 is a master TF required for definitive haematopoiesis, and more specifically is critical for EHT to occur (Cai *et al.*, 2000; Chen *et al.*, 2009; North *et al.*, 1999; Okuda *et al.*, 1996; Wang *et al.*, 1996a). *Runx1* expression begins in HE cells (North *et al.*, 1999), marking the start of, and commitment to, the haematopoietic program. Importantly, there is an enhancer 23 kb downstream (+23) of *Runx1*, which was found to be a strong driver of reporter gene expression, and to mark both HE and IACs (Bee *et al.*, 2010; Ng *et al.*, 2010; Nottingham *et al.*, 2007). This +23 enhancer is active in HE as early as E8.5, yet precedes endogenous *Runx1* transcription at this stage (Swiers *et al.*, 2013a), suggesting the existence of a pre-HE population prior to commitment to the haematopoietic program.

EHT has been well characterised as a number of specific intermediate cell states following *Runx1* expression in HE, based on various markers and transcriptional profiles. HE and IAC cells undergo maturation through a series of well-defined cell states, classified by CD41, CD43, and CD45 cell surface markers (Ottersbach, 2019; Rybtsov *et al.*, 2014; Rybtsov *et al.*, 2011; Taoudi *et al.*, 2008). HE first undergoes budding into the lumen of the dorsal aorta and differentiates into a pro-HSC state in E9.5 embryos, with a $CD41^{+}CD43^{-}$ phenotype (Rybtsov *et al.*, 2014). From E10.5 to E11.5 these pro-HSCs then mature into pre-HSC type I cells with the

expression of CD43, and subsequently pre-HSC type II cells with the expression of CD45 (Rybtsov *et al.*, 2011; Taoudi *et al.*, 2008). EHT can also be modelled in vitro, as mouse embryonic stem cells (mESCs) can be induced to differentiate into haematopoietic lineages (Carpenter *et al.*, 2011; Galic *et al.*, 2006; Lu *et al.*, 2011; Woll *et al.*, 2009), though it is important to highlight it closer represents yolk sac EHT, than dorsal aorta EHT (McGrath *et al.*, 2015).

1.3.1 Transcriptional regulation of EHT

As cells mature through the various EHT states, dynamic transcriptional changes occur (Baron *et al.*, 2018; Bergiers *et al.*, 2018; Gao *et al.*, 2020; Goode *et al.*, 2016; Hou *et al.*, 2020; Ottersbach, 2019; Solaimani Kartalaei *et al.*, 2015; Swiers *et al.*, 2013a; Vink *et al.*, 2020; Zeng *et al.*, 2019; Zhu *et al.*, 2020). One example involves the Notch signalling pathway, where studies have shown that activation of Notch TFs leads to upregulated *Gata2* and *Runx1*, and subsequent HSC expansion (Burns *et al.*, 2005; Guiu *et al.*, 2013; Robert-Moreno *et al.*, 2005). While Notch activity is required for activation of essential TFs, it must also be restricted for the progression of pre-HSC I to pre-HSC II cells (Souilhol *et al.*, 2016), highlighting a requirement for tight temporal control of gene expression. Similarly, the TF Sox17 has been found to specify HE identity, yet is considered a repressor of Runx1 and Gata2 function and must be downregulated for EHT progression (Clarke *et al.*, 2013; Lizama *et al.*, 2015; Ottersbach, 2019).

The regulation of *Runx1* in EHT has been characterised by several studies. Multiple TFs, such as Meis1, Gata2, Tal1, and Ets factors, have been identified to regulate the +23 enhancer (Nottingham *et al.*, 2007; Schütte *et al.*, 2016). Alongside Runx1, *Gata2* expression is also important for the generation of HSCs through EHT, and has a further role in HSC maintenance (Ling *et al.*, 2004; Tsai *et al.*, 1994). Runx1 has also been found to positively autoregulate itself, through interacting with the *Runx1* P1 promoter (Martinez *et al.*, 2016). Specifically within HE, *Runx1* expression has been associated with the activity of genes involved

in both cell adhesion and cell migration (Lie-A-Ling *et al.*, 2014), highlighting the potential for Runx1 to drive the morphological changes underpinning the HE budding process. Further, Runx1 is known to upregulate expression of the *Gfi1* and *Gfi1b* TFs, which subsequently leads to a Gfi1/Gfi1b mediated loss of endothelial identity (Lancrin *et al.*, 2012; Wilson *et al.*, 2010). This activity is essential for HE to undergo morphological changes, and is a requirement for haematopoietic development. Runx1 is also well characterised to activate *Spi1* (which encodes PU.1) (Huang *et al.*, 2008), an essential regulator of later haematopoietic populations that drives myeloid lineage commitment (Imperato *et al.*, 2015).

EHT involves diverse transcriptional changes across several cell type intermediates, and as Runx1 is such an important TF for this to occur, this makes for an exciting system in which to test a GRN model and investigate the role of the Runx1 TF.

1.4 *MLL*-rearranged leukaemia

Leukaemia is a cancer of the blood, of which there are multiple subtypes. Two prevalent subtypes of leukaemia are acute myeloid leukaemia (AML) and acute lymphoblastic leukaemia (ALL). These diseases can affect both young and old people alike, with ALL maintaining a particularly high prevalence in children (Fig. 1.4A, Meyer *et al.*, 2018; National Cancer Institute, 2022). Large strides have been made in curing both childhood ALL and AML, however infant leukaemias (patients $<$ 1 year of age) are still associated with a dismal prognosis (Pieters *et al.*, 2019; Rice *et al.*, 2020). A particular characteristic of infant leukaemias is the frequency of chromosomal rearrangements at the *Mixed-Lineage Leukemia* (*MLL*, also known as *KMT2A*) gene. These chromosome translocations fuse *MLL* in frame to one of a wide number of partner genes, creating novel fusion proteins (MLL-FPs). While only 10% of acute leukaemias across all ages carry *MLL* rearrangements (*MLL*r), this frequency rises to 70% in infant leukaemias (Mann *et al.*, 2010; Meyer

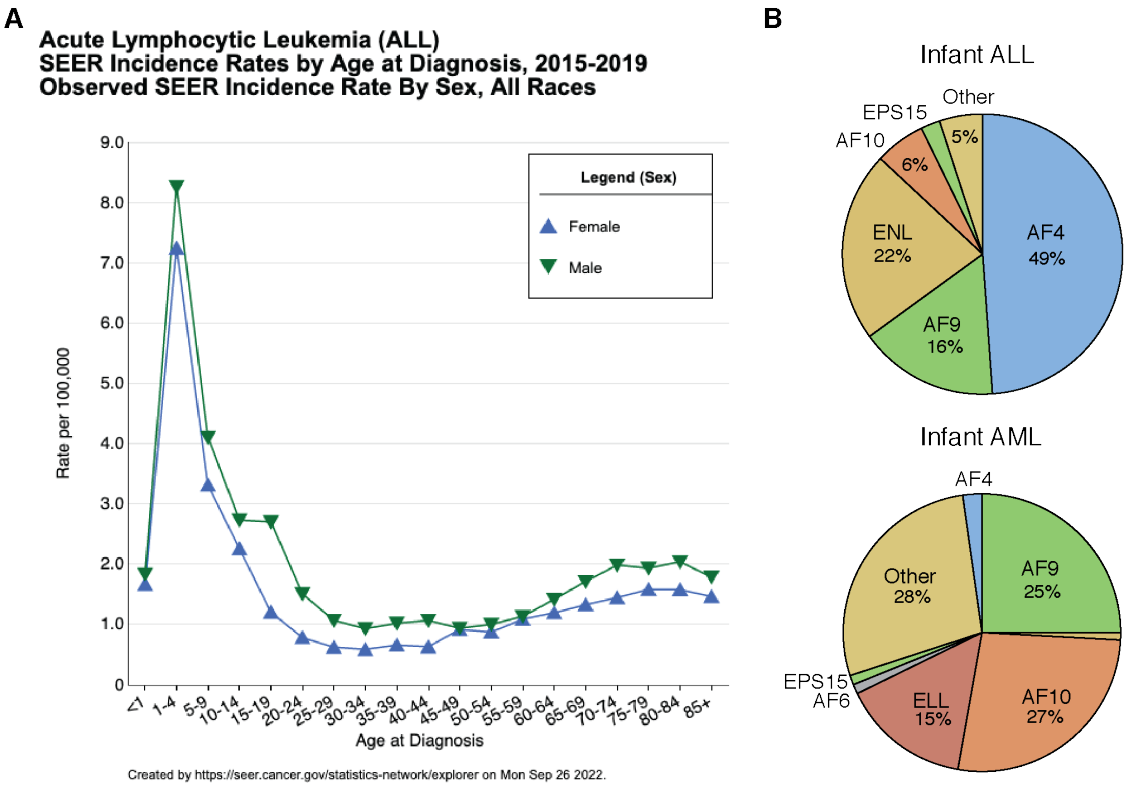


Figure 1.4: Incidence of ALL and infant *MLL*r fusion partners. (A) ALL incidence rates by age, based on data collected between 2015 and 2019. *Data and image sourced from the National Cancer Institute (NIH) Surveillance Epidemiology and End Results (SEER) programme (National Cancer Institute, 2022).* **(B)** Frequency of *MLL* fusion partners in *MLL*r infant ALL and AML. *Figure adapted from Meyer* et al.*, 2018.*

et al., 2013; Muntean *et al.*, 2012; Winters *et al.*, 2017). *MLL*r is associated with a poor prognosis, with a 20-40% 5 year event-free survival, compared to $>$ 60% in wild-type *MLL* (wt-*MLL*) (Hilden *et al.*, 2006; Tomizawa *et al.*, 2007; Winters *et al.*, 2017). Additionally, *MLL*r leukaemias respond poorly to treatment, and 50-60% of *MLL*r leukaemias in remission result in relapse (Milne, 2017; Reaman *et al.*, 1999; Winters *et al.*, 2017). There is a clear clinical need to investigate *MLL*r leukaemias, due to the prevalence and severity in infant patients, as well as its resistance to treatment strategies.

Rearrangements of the *MLL* gene can result in translocations with a wide range of partner genes, resulting in unique fusion proteins. In *MLL*r infant ALL (iALL), the most common fusion is between *MLL* and *AF4*, resulting in an MLL-AF4 fusion protein (Fig. 1.4B, Fig. 1.5A, Meyer *et al.*, 2018). Conversely, in *MLL*r

infant AML, *AF4* is a relatively rare, with *ELL*, *AF10*, and *AF9* accounting for 67% of *MLL* fusion partners. There are also cases where MLL-AF4 iALLs relapse after treatment, arising as an AML derived from the original leukaemic clone (Dorantes-Acosta *et al.*, 2012; Gardner *et al.*, 2016). This suggests that while different fusion partners are biased to specific lineages, a single fusion protein can drive both ALL and AML phenotypes.

1.4.1 Structure and function of the MLL fusion protein complex

MLL is an important developmental protein, particularly for haematopoiesis (Crump *et al.*, 2019; McMahon *et al.*, 2007). MLL is a methyltransferase that methylates histone H3 lysine 4, and this activity is driven through its SET domain, which was found to be important for wt-MLL upregulation of *Hox* genes (Milne *et al.*, 2002; Nakamura *et al.*, 2002). *MLL* translocations involve the fusion of N-terminus MLL to the C-terminus of a partner protein, resulting in loss of the SET domain and wt-MLL methyltransferase activity (Fig. 1.5B, Winters *et al.*, 2017; Zeleznik-Le *et al.*, 1994). However, the DNA binding activity of wt-MLL is mostly harboured within the N-terminus, and as such MLL binding affinities strongly direct binding of the MLL-FP (Winters *et al.*, 2017; Zeleznik-Le *et al.*, 1994). This consists of a CxxC domain, which binds unmethylated CpG (uCpG) islands (Birke *et al.*, 2002), as well as multiple AT-hooks that promote binding to AT-rich DNA (Zeleznik-Le *et al.*, 1994) (Fig. 1.5B). DNA binding is also thought to be driven through N-terminal MLL recruitment of Menin, and subsequently LEDGF, as LEDGF binds H3K36me2 marks (Hughes *et al.*, 2004; Milne *et al.*, 2005a; Yokoyama *et al.*, 2008; Zhu *et al.*, 2016), though this has been shown to be dispensable for targeting to the *HOXA9* gene (Milne *et al.*, 2010). The C-terminus of wt-MLL also contains multiple plant homeodomain (PHD) fingers and a bromodomain that contribute to chromatin binding (Fig. 1.5B), and are lost in the MLL-FP (Zeleznik-Le *et al.*, 1994).

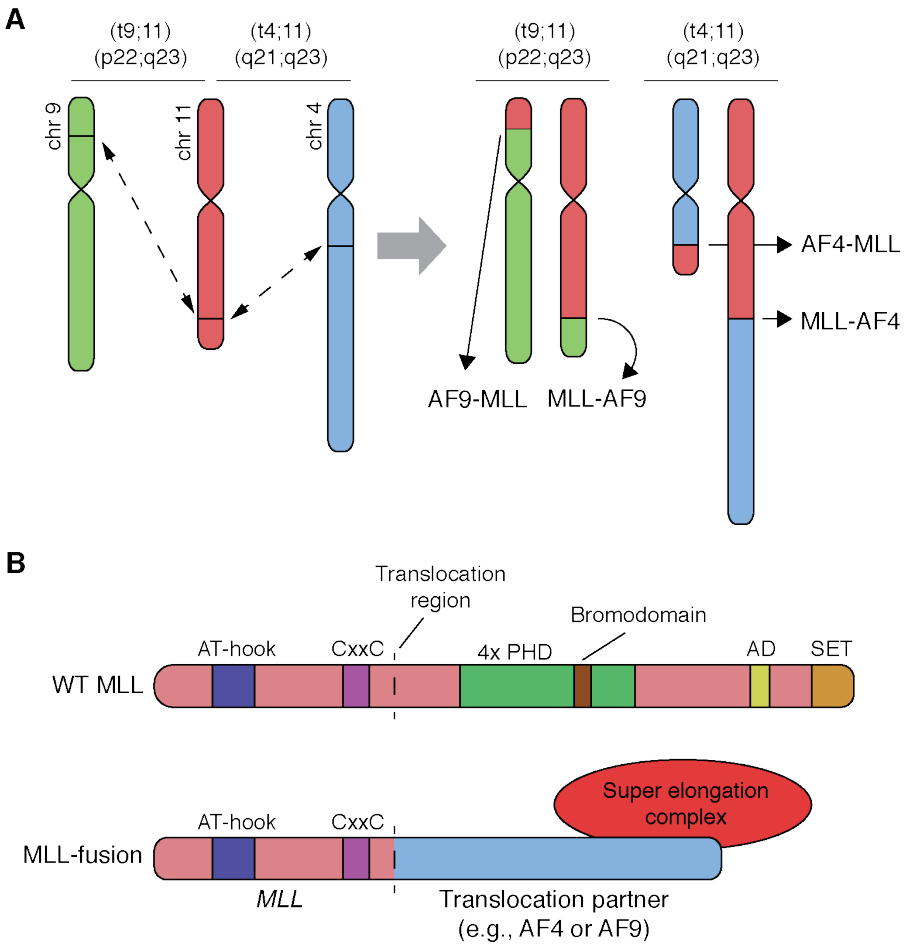


Figure 1.5: Schematic of *MLL* translocations, and wild type and MLL fusion protein structure. (A) Schematic of the chromosomal rearrangements resulting in MLL-AF4 and MLL-AF9 fusion proteins. *MLL* is located on chromosome 11, and *AF4* and *AF9* are located on chromosome 4 and 9, respectively. Horizontal black lines on chromosomes indicate the translocation break points. Dashed lines indicate translocation. **(B)** Simplified schematic of MLL fusion protein structure, adapted from Winters *et al.*, 2017.

Both MLL-AF4 and MLL-AF9 fusion proteins drive transcription through the recruitment of various factors, driven through the AF4 or AF9 C-terminus. Importantly, this includes the recruitment and assembly of a super elongation complex (SEC) that drives gene transcription (Fig. 1.5B), consisting of both AF4 and AF9, as well as AFF4, ENL, ELL1, ELL2, and ELL3, among others (Ballabio *et al.*, 2012; Lin *et al.*, 2011; Lin *et al.*, 2010; Slany, 2020; Smith *et al.*, 2011; Takahashi *et al.*, 2020; Yokoyama *et al.*, 2010). This complex contains multiple common MLL fusion partners, and as such highlights how different MLL-FPs can

have similar behaviours by recruiting a similar overall complex. The MLL-FP complex also recruits Disruptor of Telomeric Silencing 1-like (DOT1L) (Bernt *et al.*, 2011; Biswas *et al.*, 2011; Krivtsov *et al.*, 2008; Lacoste *et al.*, 2002; Lin *et al.*, 2016; Milne *et al.*, 2005b; Mueller *et al.*, 2007; Okada *et al.*, 2005; Slany, 2020), which is a methyltransferase for H3 lysine 79 (H3K79) (Steger *et al.*, 2008). H3K79 methylation has been associated with strong transcriptional overexpression at a subset MLL-FP targets, such as *HOXA9* and *RUNX1*, and perturbation of DOT1L activity significantly downregulates gene expression (Godfrey *et al.*, 2019; Kerry *et al.*, 2017; Milne *et al.*, 2005b; Wilkinson *et al.*, 2013). Additionally, in *MLL*r leukaemia H3K79me is also found enriched at a subset of intragenic enhancers, and maintains enhancer-promoter (E-P) interactions (Godfrey *et al.*, 2019). A related cofactor that is associated with the MLL-FP complex, though not directly recruited, is BRD4. BRD4 binds acetylated chromatin via its bromodomain (Dey *et al.*, 2003), but in addition to this it was found that BRD4 was associated with members of the SEC, and further that BRD4 was a potential therapeutic target for MLL-AF9 leukaemias (Dawson *et al.*, 2011; Zuber *et al.*, 2011b). BRD4 directly interacts with P-TEFb and stimulates transcription by increasing P-TEFb dependent Pol II phosphorylation (Jang *et al.*, 2005; Yang *et al.*, 2005). It has also been reported that DOT1L and BRD4 functionally cooperate in MLL leukaemias, through EP300 acetylating H4K5 at methylated H3K79 regions, and subsequent BRD4 recruitment to the acetylated chromatin (Gilan *et al.*, 2016).

An interesting feature of MLL-FPs, most characterised with MLL-AF4, is the observation that at a subset of gene targets MLL-AF4 binds to uCpG promoters then spreads into the gene body (Kerry *et al.*, 2017). These "spreading targets" are highly overexpressed, and spreading is associated with enriched Menin and H3K79 methylation. Interestingly, these overexpressed targets were also highly susceptible to DOT1L inhibition, highlighting DOT1L as key driver at this subset of targets.

1.4.2 Transcriptional impact of MLL fusion proteins

As MLL-AF4 and MLL-AF9 recruit a SEC, and are associated with transcriptional coactivators (i.e., BRD4), these fusion proteins are considered to act as drivers of gene transcription rather than repressors. Interestingly, *MLL*r leukaemias are associated with few or no cooperating mutations (Andersson *et al.*, 2015; Bardini *et al.*, 2011; Bardini *et al.*, 2010; The Cancer Genome Atlas Research Network, 2013), suggesting that the MLL fusion protein alone is sufficient to drive leukaemogenesis. In fact, numerous studies have shown that expression of MLL-AF4 or MLL-AF9 is sufficient to induce leukaemia (Krivtsov *et al.*, 2008; Krivtsov *et al.*, 2013; Rice *et al.*, 2021). As MLL fusion proteins function by overexpressing genes, and are sufficient on their own for leukaemogenesis, *MLL*r leukaemias may be considered to be an entirely transcriptionally driven disease. This makes for an ideal system to probe aberrant gene expression profiles with a GRN approach.

BCL2 is an MLL-AF4 and MLL-AF9 target of particular importance. It has been reported that not only does MLL-AF4 bind to the *BCL2* promoter, but also a downstream enhancer, resulting in overexpression of the gene (Benito *et al.*, 2015; Godfrey *et al.*, 2017). MLL-AF4 leukaemias are highly sensitive to *BCL2* perturbation (Robinson *et al.*, 2008), in keeping with the role of BCL-2 in deterring apoptosis (Czabotar *et al.*, 2014; Singh *et al.*, 2019). This highlighted *BCL2* as a key gene in cell survival. As such, venetoclax, a drug targeting BCL-2 protein, has been a successful treatment of MLL-AF4 leukaemias (Benito *et al.*, 2015; Khaw *et al.*, 2016; Niu *et al.*, 2014).

MLL-AF4 and MLL-AF9 have also been found to overexpress *MYC*, a TF and oncogene that drives proliferation (Dang, 2012; Prange *et al.*, 2017). *MYC* overexpression typically results in apoptosis, either through indirectly activating p53 (p53-dependent) or modulating the balance between pro-apoptosis and anti-apoptotic factors (p53-independent) (Fairlie *et al.*, 2021). However, the overexpression of

BCL2 prevents activation of the p53-independent apoptotic pathway, allowing MLL-AF4 and MLL-AF9 to upregulate *MYC* and increase proliferation. (Bissonnette *et al.*, 1992; Fairlie *et al.*, 2021; Fanidi *et al.*, 1992). As such, both *BCL2* and *MYC* must be expressed for the full leukaemic potential of *MLL*r leukaemias.

Another key gene is *PROM1* (encoding CD133 protein), which is an MLL-AF4 spreading target that is important for leukaemic growth (Godfrey *et al.*, 2021; Kerry *et al.*, 2017; Mak *et al.*, 2012). Interestingly, it was found that not only does MLL-AF4 drive expression through the promoter, but H3K79 methylation deposited by DOT1L at an intragenic enhancer was found to maintain E-P interactions, potentially enhancing promoter activity (Godfrey *et al.*, 2021). Further, this intragenic enhancer also interacted with a nearby gene *TAPT1*, highlighting the capacity for complex regulation driven by MLL-AF4, not just through promoter binding, but modulation of enhancer activity and E-P interactions.

RUNX1 and MYB are two TFs of particular interest. *RUNX1* is overexpressed in MLL-AF4 ALL leukaemia, and is an MLL-AF4 spreading target (Kerry *et al.*, 2017; Wilkinson *et al.*, 2013). It has also been found that *RUNX1* perturbation is detrimental for leukaemic growth (Wilkinson *et al.*, 2013). While RUNX1 itself is a potent transcriptional regulator, Wilkinson *et al.*, 2013 suggested an additional regulatory mechanism whereby RUNX1 interacts with the AF4-MLL (the reciprocal fusion protein of MLL-AF4) complex. MYB is a known to be upregulated by MLL-AF9 in MLL-AF9 AML, after which these AML become addicted to MYB (Zuber *et al.*, 2011a). Both RUNX1 and MYB are expressed in numerous haematopoietic populations, in addition to the critical role of RUNX1 in EHT (section 1.3), and are important regulators of several lineages (Ichikawa *et al.*, 2013; Wang *et al.*, 2018). Probing the activity of these deregulated TFs in leukaemia will be important to fully understand the regulation of these diseases, as well as how aberrantly expressed transcriptional networks contribute to leukaemogenesis.

1.5 Thesis overview and aims

In this thesis I have focused on the application of GRN models to investigate gene regulation in a normal developmental process, and gene regulation in aberrantly regulated processes, most notably leukaemia. Gene regulation is complex to model in vivo, and TFs add a layer of complexity that is difficult to address. The specific combination of TFs, not just expressed in a cell but bound at a specific genomic locus, can drastically alter regulatory output. This specific combinatorial code underpins all developmental processes, can drive change in cell identity, and can be a critical component for understanding how diseases are driven.

Haematopoiesis involves complex gene regulation and multiple lineages, and cell identity is relatively plastic, with cells capable of being reprogrammed into different lineages (Batta *et al.*, 2014; Riddell *et al.*, 2014; Schütte *et al.*, 2016). As a result, it is also an active field of research for applying GRN models (Rothenberg *et al.*, 2021). EHT is a particularly interesting process associated with the developmental origin of HSCs, that requires dose-specific and temporally controlled activation of TFs and dynamic regulation of gene expression. However, the disruption of normal regulatory processes can result in disease, as seen with *MLL*r acute leukaemias. MLL-AF4 and MLL-AF9 leukaemias are interesting haematopoietic malignancies, that are unique in being driven by a single translocation event and are a transcriptional disease (Andersson *et al.*, 2015; Bardini *et al.*, 2011; Bardini *et al.*, 2010; Krivtsov *et al.*, 2008; Krivtsov *et al.*, 2013; Rice *et al.*, 2021; The Cancer Genome Atlas Research Network, 2013).

The well-defined populations of EHT and dynamic gene regulation make for a promising system to probe using GRN approaches. As the RUNX1 TF is critically required for EHT, it is a pivotal TF to investigate in this GRN. *RUNX1* is also an important MLL-AF4 target overexpressed in *MLL*r ALL, and as such *MLL*r leukaemias represent a system in which RUNX1 behaviour can be studied within a misregulated, leukaemic GRN. Studying GRN models of these systems offers an

interesting view of RUNX1 behaviour in different contexts: normal and aberrant haematopoiesis. MYB is also an interesting TF that, while not critical for EHT, is an essential overexpressed gene MLL-AF9 AML. Studying GRN models with a view of RUNX1 and MYB behaviour, with a particular focus on how these TFs cooperate with other regulators, will offer interesting comparisons.

There are three questions central to this thesis project:

1. What novel regulatory mechanisms underpinning key processes can be understood using GRN models?

2. What insight into TF combinatorial logic and regulatory synergy driving processes do these GRNs offer?

3. How are normal transcriptional networks deregulated in *MLL*r leukaemia?

I have established and explored a model of EHT regulation in chapter 3 to investigate key regulatory interactions involving Runx1, and to probe cases of TF cooperation that may underlie the EHT process. Further, in this chapter I have explored the EHT in relation to the pre-HE population, and identified genes characterising this population, and TFs that may drive activity of the +23 *Runx1* enhancer at this early stage. I have also generated a model of MLL-AF4 and RUNX1 leukaemia in chapter 4, where I have explored key drivers of MLL-AF4 ALL, and the specific cooperation between MLL-AF4 and RUNX1, as well as RUNX1 and MYB. This work was extended in chapter 5, where I have investigated the leukaemic transformation of a normal granulocyte-monocyte progenitor (GMP) GRN after inducing *MLL-AF9* expression. This final chapter attempts to explore how a normal GRN can be transformed over time, following induction of leukaemia.

This study serves to further our knowledge of the regulation of EHT, particularly through combinatorial logic. It also improves our understanding of the pre-HE population, and offers insight into the potential early regulators of the +23 *Runx1* enhancer. Through studying MLL-AF4 and MLL-AF9 leukaemias, this thesis

highlights concepts characterising the perturbation of normal GRNs in *MLL*r leukaemia, driven in part through TF cooperation.

2

Materials and methods

Contents

2.1 Mouse husbandry 30
2.2 Embryo studies 30
2.2.1 Timed matings and embryo collection 30
2.2.2 Embryo tissue dissection and cell sorting 31
2.2.3 Whole mount immunofluorescence imaging 32
2.3 Cell culture 33
2.3.1 Culturing mESCs 33
2.3.2 Culturing THP1, SEM, and RS4;11 cell lines 34
2.3.3 mESC differentiations 34
2.4 MLL-AF9 transduced GMP experiments 35
2.4.1 Bone marrow GMP isolation 35
2.4.2 Transduction of mouse GMPs 36
2.4.3 MethoCult expansion of transduced GMPs 37
2.5 siRNA knockdowns 38
2.6 Colony-forming assays 38
2.7 Drug treatments and dose-response curve analysis . . 39
2.8 Western blot analysis 39
2.9 Gene expression analysis by qRT-PCR 40
2.9.1 Total RNA extraction and cDNA synthesis 40
2.9.2 qRT-PCR analysis 41
2.10 RNA-seq methods 42
2.10.1 Poly-A RNA-seq library preparation 42
2.10.2 Nascent RNA-seq library preparation 43
2.10.3 Library preparation from embryonic populations 43
2.11 Chromatin immunoprecipitation (ChIP) methods . . . 44
2.11.1 ChIP 44
2.11.2 ChIP qPCR 45
2.11.3 ChIP-seq library preparation 46
2.11.4 ChIPmentation library preparation 47
2.12 ATAC-seq library preparation 48
2.13 RNA+ATAC multiome library preparation 49
2.14 Computational analysis 50

2.14.1 RNA-seq analysis 50
2.14.2 ChIP-seq and ATAC-seq analysis 52
2.14.3 Dimensionality reduction methods 53
2.14.4 Single-cell sequencing analysis 53
2.14.5 Pseudotime analysis 55
2.14.6 Pathway enrichment 56
2.15 GRN construction and analysis 56
2.15.1 TF binding motif analysis 56
2.15.2 EHT GRN methodology 56
2.15.3 MLL-AF4 leukaemia GRN methodology 57
2.15.4 MLL-AF9 GMP leukaemogenesis GRN methodology . . 58
2.15.5 Centrality calculation 59
2.15.6 TF Co-interaction analysis 60
2.16 Statistics 60
2.17 Sourcing of patient samples 60
2.18 Sourcing of published datasets 61

2.1 Mouse husbandry

All procedures involving mice were in compliance with UK Home Office regulations and the Oxford University Clinical Medicine Ethical Review Committee. Mice were housed in individually ventilated cages with free access to food and water and maintained in a 12-hour light-dark cycle.

2.2 Embryo studies

2.2.1 Timed matings and embryo collection

Timed pregnancies were generated through overnight mating. To acquire transgenic embryos, matings were set up with (CBA × C57BL/6)/F1 females and 23GFP (*Runx1* +23 GFP enhancer reporter) transgenic males (maintained on a mixed (CBA × C57BL/6) background). To acquire wild type embryos, matings were set up with (CBA × C57BL/6)/F1 females and (CBA × C57BL/6)/F1 males. Confirmation of a vaginal plug after overnight mating was marked embryonic day

(E)0.5. At the appropriate developmental time point, mice were killed by a Schedule 1 method and embryos extracted from the uteri, and staged by somite pairs.

2.2.2 Embryo tissue dissection and cell sorting

Embryo dissections for RNA-seq and ATAC-seq experiments (chapter 3) were performed as previously described (Swiers *et al.*, 2013a) by members of the de Bruijn lab. Para-aortic splanchnopleura (PAS)/AGM and VU vessels were dissected from E8.5 (PAS only), E9.5 and E10.5 embryos in PBS supplemented with 10% fetal calf serum (FCS), 50 U/ml penicillin, and 50 μg/ml streptomycin (Gibco). PAS/AGM+VU tissues were pooled and incubated for 15 min at 37°C in PBS without calcium or magnesium, containing 10% FCS, 50 U/ml penicillin, 50 μg/ml streptomycin, and 0.1% (w/v) type I collagenase (Sigma). Cells were disassociated by gentle pipetting.

Cells were washed, resuspended, and incubated in an antibody cocktail (Table 2.1) optimally diluted in 10% FCS/PBS. Cells were incubated for 30 min 4°C, washed, and resuspended in 10% FCS/PBS with 1/1000 Hoechst 33258 (Thermo fisher) to stain dead cells. Haematopoietic populations were isolated by fluorescence-activated cell sorting (FACS) on a BD FACSAria Fusion. Sorted phenotypes include E8.5 endothelial cells (EC, Ter119$^-$VECad$^+$CD41$^-$CD45$^-$23GFP$^-$) and pre-HE (Ter119$^-$VECad$^+$CD41$^-$CD45$^-$23GFP$^+$), E9.5 HE (Ter119$^-$VECad$^+$-CD41$^-$CD45$^-$23GFP$^+$) and pro-HSC/haematopoietic progenitors (Ter119$^-$VECad$^+$CD41$^+$CD45$^-$), and E10.5 pre-HSC I (Ter119$^-$VECad$^+$-CD41$^+$CD45$^-$) and pre-HSC II (Ter119$^-$VECad$^+$CD41$^+$CD45$^+$). Instruments were set up using unstained and single stain controls and sort gates were drawn according to fluorescent minus one (FMO) controls. All sorts were done by the staff of the WIMM Flow Cytometry core facility.

Table 2.1: Antibodies used for flow cytometry analysis and FACS sorting of EHT cell populations.

Target	Conjugate	Dilution	Catalogue number	Company
CD41	PE	1:400	558040	Pharmigen
CD45	APCef780	1:200	13-5821-81	eBioscience
Ter119	PE-Cy7	1:200	557853	Pharmigen
VE-cadherin	APC	1:200	17-1441-82	eBioscience

All antibodies reactive in mouse, and raised in rat.

2.2.3 Whole mount immunofluorescence imaging

Embryos at day E8.5 were processed for whole-mount immunofluorescence analysis as described in Yokomizo *et al.*, 2012. Embryos were fixed in 4% paraformaldehyde for 30 min 4°C, then washed with PBS 3x 10 min each. Embryos were dehydrated by 1x 50% methanol/PBS 10 min, followed by 2x 100% methanol 10 min washes. Dehydrated embryos were stored at -20°C.

Before antibody staining, embryos were rehydrated by 1x 50% methanol/PBS 10 min, followed by 2x PBS 10 min washes. Embryos were kept in the dark during all wash and incubation steps. Embryos were incubated with blocking buffer (0.4% triton X-100, 10% FCS in PBS) for 3 hours 4°C to block non-specific binding of antibodies. Primary antibodies were diluted in blocking buffer (see table 2.2), and embryos were incubated with 200 μl antibody mix overnight at 4°C with rocking. Embryos were washed 3x with blocking buffer for 2 hours each wash. Secondary antibodies were diluted in blocking buffer (see table 2.2), and embryos were incubated with 200 μl antibody mix overnight at 4°C with rocking. Embryos were washed 3x with blocking buffer for 2 hours each wash. Embryos were washed 3x 20 min with PBS with 0.4% triton-X 100. Embryos were dehydrated in methanol, as above. Using molten 1% agarose, embryos were anchored onto a coverslip with a FastWell reagent barrier (664113-GRA, Grace Bio-Labs) sealed, and the FastWell was filled with methanol. Embryos were washed 3x 1 min with methanol. A 1:1 solution of benzyl alcohol (#402834, Sigma) and benzyl benzoate (#0215483990,

MP Biomedicals) was prepared (BABB solution). Embryos were washed 3x 1 min with a 50% methanol and 50% BABB solution. 100% BABB solution washes were performed until the embryos were sufficiently clear. A coverslip was applied to the surface of the FastWell, and sealed with nail polish to prevent leakage.

Image acquisition was performed using a Zeiss AXIO Observer.Z1 inverted microscope equipped with a Zeiss LSM-880 confocal system using an 25x LDLCI PlnApo NA:0.8 DI objective. Image acquisition settings were adjusted on the basis of unstained, single stain, and FMO controls.

Table 2.2: Antibodies used for whole-mount immunofluorescence.

Target	Conjugate	Dilution	Raised in	Catalogue number	Company
CD31	Unconjugated	1:200	Goat	AF3628	R&D Systems
GFP	Unconjugated	1:200	Chicken	ab13970	Abcam
CD109	Unconjugated	1:100	Rabbit	24765S	Cell Signaling
anti-Goat IgG	AF647	1:200	Donkey	A21447	Invitrogen
anti-Chicken IgG	AF488	1:200	Donkey	703-545-155	Jackson
anti-Rabbit IgG	AF594	1:200	Donkey	A21207	Invitrogen

2.3 Cell culture

All cell lines were grown in a 37°C humidified incubator with 5% CO_2, and were confirmed to be free of mycoplasma.

2.3.1 Culturing mESCs

Mouse embryonic stem cells (mESCs) (E14, strain 129/Ola) were cultured in G-MEM medium (Thermo Scientific) with 10% FCS (Gibco), 100 mM non-essential amino acids (Gibco), 2 mM L-glutamine (Gibco), 100 mM sodium pyruvate (Gibco), 100 μM β-mercaptoethanol (Gibco), and leukaemia inhibitor factor (LIF, prepared in house). mESCs were passaged every 2-3 days with 0.05% Trypsin (Gibco).

2.3.2 Culturing THP1, SEM, and RS4;11 cell lines

SEM cells, a MLL-AF4 B cell ALL line (Greil *et al.*, 1994), were purchased from DSMZ (www.dsmz.de), and cultured in Iscove's modified Dulbecco's medium (IMDM) with 10% FCS and 1× GlutaMAX. Cell density was maintained between 5 x 10^5/ml and 2 x 10^6/ml. RS4;11 and THP-1 cells were purchased from ATCC (www.atcc.org) and cultured in RPMI-1640 with 10% FBS and 1× GlutaMAX. Cell density was maintained between 5 x 10^5/ml and 1.5 x 10^6/ml.

2.3.3 mESC differentiations

Serum free mESC differentiations were performed by L. Greder, based on Pearson *et al.*, 2015 with minor modifications, for validation experiments in section 3.6.1. mESC growth media was replaced with media -LIF for 24 hours. 7.5 x 10^5 mESCs were resuspended in 20 ml StemPro-34 with StemPro-Nutrient supplement (Gibco) containing 8 mg/ml ascorbic acid (Sigma), 100 U/ml transferrin (Roche), 2 mM L-glutamine (Gibco), and 5 ng/ml BMP-4 (R&D Systems), then cultured for 72 hours (day 3) to form embryoid bodies (EBs). Medium was supplemented with 5 ng/ml bFGF (R&D Systems) and 5 ng/ml Activin A (Peprotech). After 24 hours (day 4), EBs were disassociated in 0.05% Trypsin (Gibco) and stained using anti-Flk1 APC conjugated antibody (17-5821-81, eBioscience) diluted 1/100 in 5% FCS/PBS for 10 min 4°C. Flk1+ cells were isolated by FACS using a BD FACSAria Fusion. 6 x 10^5 Flk1+ cells were plated in differentiation medium with 5 ng/ml BMP-4 (R&D Systems), 5 ng/ml Activin A (Peprotech), 10 ng/ml SCF (Peprotech), and 25 ng/ml VEGF (Peprotech). After 24 hours (day 5), medium was refreshed containing 10 ng/ml SCF (Peprotech) and 25 ng/ml VEGF (Peprotech). After 48 hours' culture (day 7), cells were harvested for flow cytometry analysis. Suspension and adherent cells were disassociated by 0.05% Trypsin (Gibco). Up to 1 x 10^6 cells were washed in 5% FCS/PBS, then resuspended in 50 μl antibody cocktail (Table 2.1) and incubated for 10 min 4°C. Cells were washed and resuspended in 5% FCS/PBS with 1/1000 Hoechst 33258 (Thermo fisher) to stain dead cells, before

analysis by flow cytometry. Instruments were set up using unstained and single stain controls. Sort gates were drawn according to FMO controls. All sorts were done by the staff of the WIMM Flow Cytometry core facility.

2.4 MLL-AF9 transduced GMP experiments

2.4.1 Bone marrow GMP isolation

Granulocyte-Monocyte Progenitors (GMPs) were isolated from mouse bone marrow (BM), with assistance from S. Valletta and S. Rice (Roy lab). GMPs were harvest from C57BL/6 *Myb*$^{+/+}$ and *Myb*$^{\mathrm{MRE/MRE}}$ mice (described in section 5.1). *Myb*$^{\mathrm{MRE/MRE}}$ mice were generated by insertion of a *miR-196b* miRNA response element (MRE) into the 3' UTR of *Myb*, as generated in (Lau, 2022), and were kindly provided by I. Lau. Female mice between 8 and 11 weeks of age (age matched between genotypes) were killed by a Schedule 1 method, and tibiae and femurs were dissected out. Tibiae and femurs from 3-5 mice (per replicate, per genotype) were pooled and crushed using a pestle and mortar, then cells were passed through a 50 μM filter with 10 ml 5% FCS/PBS. Cell counts were obtained using a Pentra ES 60 (Horiba). Cells were enriched for c-Kit expression by resuspension with 2.5 μl magnetic c-Kit beads (#130-091-224, Miltenyi) in 100 μl 5% FCS/PBS per 1 x 10^8 cells, and incubation at 4°C 20 min with gentle agitation. Cells were washed, then c-Kit+ cells were isolated using MACS columns, as per manufacturer's instructions, and counted using a Pentra ES 60 (Horiba).

c-Kit+ cells were incubated with either PE-Cy7 conjugated anti-CD16/32 monoclonal antibody (25-0161-82, eBioscience), or unconjugated anti-CD16/32 monoclonal antibody (14-0161-81, eBioscience) for controls, to block non-specific antibody binding at 4°C 15 min. Antibody cocktails for controls and samples were prepared at 2X concentration, and added to cells incubated in CD16/32 antibodies at a 1:1 volume for a final 1X concentration, and incubated 4°C 15 min. Antibody panel details are in table 2.3. Cells were washed and resuspended

in 5% FCS/PBS with 1X 7-aminoactinomycin D (7AAD) viability dye (#A9400, Sigma Aldrich). GMPs were then isolated by FACS on a BD FACSAria Fusion, as a $Lin^{-}cKit^{+}CD41^{-}CD150^{-}CD16/32^{+}$ phenotype. Instruments were set up using unstained and single stain controls. Sort gates were drawn according to FMO controls. All sorts were done by the staff of the WIMM Flow Cytometry core facility.

Isolated GMPs were cultured overnight in 500 μl OptiMEM with 20% FBS, 50 U/ml penicillin, 50 μg/ml streptomycin, and cytokines (10 ng/μl mSCF, 5 ng/μl mIL3, and 5ng/μl IL6, Stemcell technologies) in a 37°C humidified incubator with 5% CO_2.

Table 2.3: Antibodies used for GMP isolation.

Target (mouse)	Conjugate	Final dilution	Clone	Catalogue number	Company	Notes
CD16/32	Unconjugated	1/50	RA3-6B2	14-0161-81	eBioscience	Fc block
CD16/32	PE-Cy7	1/150	93	25-0161-82	Invitrogen	
CD150	APC	1/100	TC15-12F12.2	115910	BioLegend	
c-Kit	APCef780	1/1200	28B	47-1171-82	eBioscience	
CD41	BV421	1/200	MWReg30	133911	BioLegend	
Sca1	BV786	1/200	D7	563991	BD Biosciences	
CD4	PE-Cy5	1/1600	RM4-5	100514	BioLegend	Lineage
CD8a	PE-Cy5	1/1600	53-6.7	15-0081-83	eBioscience	Lineage
CD5	PE-Cy5	1/1200	53-7.3	100610	BioLegend	Lineage
Gr1	PE-Cy5	1/400	RB6-8C5	15-5931-82	Invitrogen	Lineage
Mac1	PE-Cy5	1/800	M1/70	15-0112-82	eBioscience	Lineage
B220	PE-Cy5	1/400	RA3-6B2	103210	BioLegend	Lineage

All antibodies reactive in mouse, and raised in rat.

2.4.2 Transduction of mouse GMPs

Isolated mouse GMPs were transduced using an MLL-AF9/mCherry plasmid generated by I. Lau (Lau, 2022) (see Appendix A.2 for plasmid map). This plasmid contained human *MLL-AF9* under the murine stem cell virus (MSCV) promoter, and *mCherry* under the 3-phosphoglycerate kinase (PGK) promoter, allowing constitutive activation in GMPs. To generate retroviral particles, 293T cells were co-transfected with the MSCV MLL-AF9/mCherry construct and Eco-Pak packaging plasmid (AddGene #12371), using PEIpro reagent (Polyplus) according to manufacturer's instructions. Viral supernatant was harvested after 24 and 48 hours post transfection and concentrated using Retro-X reagent (#631456, Takara).

1 part Retro-X reagent was added to 3 parts supernatant and mixed by gentle inversion, then incubated 4°C overnight. Retroviral supernatant was centrifuged 1500 RCF 4°C for 45 min, and the pellet resuspended at $1/100^{th}$ the original volume, and stored at -80°C. 24 hours after GMP isolation and culture in 500 μl OptiMEM with 20% FBS, 50 U/ml penicillin, 50 μg/ml streptomycin, and cytokines (10 ng/μl mSCF, 5 ng/μl mIL3, and 5 ng/μl IL6, Stemcell technologies) (section 2.4.1), 50 μl concentrated retrovirus and 7 μg/ml Polybrene was added to cell suspension. Cells were spinoculated by centrifugation at 1500 RCF 37°C for 1 hour, then incubated at 37°C with 5% CO_2 for 24 hours.

2.4.3 MethoCult expansion of transduced GMPs

Following GMP infection with the MSCV MLL-AF9/mCherry plasmid (section 2.4.2), 3000-7000 mCherry+ cells were sorted by FACS on a BD FACSAria Fusion directly into 1.5 ml MethoCult semi-solid methylcellulose-based media (M3434, STEMCELL technologies). Cells were vortexed in MethoCult to sufficiently distribute cells, and 1 ml was dispensed into a dish. Cells were cultured at 37°C for 7 days (week 1, W1). At W1, cells were harvested by dissolving MethoCult in IMDM + 10% FCS and washing. Cells were taken for RNA-seq (section 2.10.1), ATAC-seq (section 2.12) and ChIPmentation (section 2.11.4) experiments. 1 x 10^4 of remaining cells were added to 4 ml MethoCult (M3434) and replated as above into 3 dishes (1 ml, 2500 cells, per dish), and incubated at 37°C for 7 days. Cell harvesting and replating was performed twice more as above, each after 7 days incubation at time points week 2 (W2) and week 3 (W3). Cells harvested at W3 were validated for *MLL-AF9* expression and $Myb^{MRE/MRE}$ genotype from RNA samples by RT-PCR (cDNA generated as in section 2.9.1). RT-PCR was performed using primers spanning the human *MLL-AF9* fusion break point (5'-GAGGATCCTGCCCCAAAGAAAAG-3' & 5'-TCACGATCTGCTGCAGAATGTGTCT-3') or primers flanking the inserted *miR-196b* MRE (5'-GGGAGATTTGTGTTGTTTATGTCA-3' & 5'-GGCAGAAACTGGCTGTTGA-3'). PCR reactions were performed with the following reaction mix: 2 μl cDNA, 1 μl 10 μM forward primer, 1 μl 10 μM

reverse primer, 15 μl 2x KAPA HiFi HotStart (KK2601, Roche), 15 μl H_2O. PCR amplification was performed in a thermocycler with the following conditions: 95°C 5 min; 30 cycles of (98°C 20 s, 65°C 15 s; 72°C 30 s); 72°C 5 min; 4°C hold.

2.5 siRNA knockdowns

siRNA knock-downs (KDs) in chapter 4 were performed using a rectangle pulse EPI 2500 electroporator (Fischer, Heidelberg). 1-7 x 10^7 SEM or RS4;11 cells were mixed with 10 μl 20 μM siRNA in 500 μl media, and electroporated at 330V for 10 msec in a 2 mm cuvette. Cells were incubated at room temperature for 15 min, then added to growth media at a cell density of 1 x 10^6/ml. For 96 hours' KD experiments, cells were re-transfected as above at 48 hours post initial transfection. SEM MLL-AF4 siRNA (siMA6), RS4;11 MLL-AF4 siRNA (siMARS) and scrambled control (siMM) siRNA sequences were sourced from Thomas *et al.*, 2005. Commercial siRNA used are detailed in Table 2.4.

Table 2.4: Commercial siRNA used for perturbation experiments.

Target	Catalogue Number	Brand
RUNX1	J-003926-05	Dharmacon ON-TARGETplus single siRNA
MAZ	s8543	Ambion Silencer Select
MYB	s9110	Ambion Silencer Select
Non-targeting	D-001810-10-20	Dharmacon ON-TARGETplus non-targeting pool

2.6 Colony-forming assays

SEM cells were transfected with siRNA (as in section 2.5), and 48 hours later re-transfected with the same siRNA. 24 hours later (72 hours' KD) 500 cells were added to 4 ml IMDM MethoCult media + 20% FCS (H4100 STEMCELL technologies). Cells were vortexed to distribute them evenly, and MethoCult was dispensed into three dishes (1 ml per dish, 125 cells). Colonies were incubated for 14 days (37°C, 5% CO2) prior to counting.

2.7 Drug treatments and dose-response curve analysis

The small molecule RUNX1 inhibitor (AI-10-104) and inactive control (AI-4-88) were gifted by the Bushweller lab (Illendula *et al.*, 2016; Illendula *et al.*, 2015). In single treatment experiments, SEM cells were cultured for 24 hours in the presence of 10 μM AI-10-104, AI-4-88, or DMSO, then characterised by ChIP and qRT-PCR assays. In double treatment experiments, SEM cells were cultured for 24 hours with 10 μM AI-10-104, AI-4-88, or DMSO, as well as a range of venetoclax (S8048-SEL, Seleck) concentrations up to 10 μM. Cell viability following RUNX1 inhibitor and venetoclax co-treatment was assayed by CellTiter-Glo (CTG) luminescence assay (#G7570, Promega) according to manufacturer's instructions. 100 μl cell suspension was mixed with 100 μl CTG reagent in white walled 96 well assay plates (Thermo). Plates were loaded on a BMG FLUOstar OPTIMA plate reader, and cells lysed by 2 min orbital shaking and room temperature incubation for 10 min, then luminescence was recorded with integration time set to 0.25-1 sec. Dose-response curves were fit to the data and IC50 values calculated using the R package drc (v3.0-1, Ritz *et al.*, 2015).

2.8 Western blot analysis

Proteins were extracted from 1 x 10^6 cells by resuspension in BC300 lysis buffer (20 mM Tris-HCl pH 8, 20% glycerol, 300 mM KCl, 5 mM EDTA) + 0.5% IGEPAL, with protease inhibitors. Cells were incubated 4°C 30 min. Lysed cells were spun at 20000 RCF 4°C for 10 min, and supernatant containing protein was collected. 1x NuPAGE LDS sample buffer (#NP0008, Novex) + 1% β-mercaptoethanol was added to protein extract, then heated at 95°C for 5 min. Protein extract from the equivalent of 1 x 10^5 cells was loaded on NuPAGE 4%-12% BisTris gels (Life Technologies) and run at 180v for 1 hour in 1X MOPS SDS Running buffer (#NP0001, Invitrogen). Proteins were transferred onto polyvinylidene fluoride membrane (#IPVH00010,

Immobilon). Membranes were washed in 100% methanol, then dried. Membranes were then rehydrated and probed at 4°C overnight using antibodies diluted in 5% milk. All antibodies used were raised in rabbit, and diluted as in table 2.5. Following 3x 5 min washes with TBST (0.1% Tween 20 in tris buffered saline), the membranes were incubated with secondary antibody (goat anti-rabbit HRP, A6667, Sigma) at 1:10000 dilution in 5% milk for 1 hour at room temperature. Following 3x 5 min washes with TBST, membranes were probed using ECL Prime Western Blotting Detection Reagent (#RPN2232, Cytiva) for chemiluminescence detection. Membranes were subsequently imaged using ChemiDoc MP (BioRad).

Table 2.5: Antibodies used for western blot analysis

Target	Dilution	Catalogue number	Company
RUNX1	1/5,000	4334S	Cell Signaling
Caspase-9	1/10,000	ab202068	Abcam
GAPDH	1/10,000	A300-641A	Bethyl
Vinculin	1/50,000	ab129002	Abcam

2.9 Gene expression analysis by qRT-PCR

2.9.1 Total RNA extraction and cDNA synthesis

Total RNA was extracted from 1-5 x 10^6 SEM cells using the RNeasy Mini kit (#74104, Qiagen). On-column DNA digestion was performed as according to the RNeasy Mini protocol using DNase I (#79254 Qiagen).

RNA concentrations were determined by NanoDrop (Thermo scientific), and samples within the same experiment were diluted to the same concentration. To generate cDNA, a reaction mix of 1 μg RNA in 8 μl H_2O, 1 μl 10 mM dNTP mix (#18427-013, Invitrogen), and 1 μl 50 μM random hexamer primers (#S06405, Roche) was incubated at 65°C 5 min, and cooled to 4°C for 2 min in a thermocycler (BioRad). Reverse transcription was performed by adding 2 μl H_2O, 4 μl 5X First strand buffer (#02321, Invitrogen), 2 μl 0.1 M dithiothreitol (DTT, #00147,

Invitrogen), 1 μl 40 U/μl RNaseOUT (#51535, Invitrogen), and 1 μl 200 U/μl SuperScript III Reverse Transcriptase (#56575, Invitrogen), and incubated in a thermocycler for: 20°C 2 min; 50°C 1 hour; 75°C 20 min. cDNA product was then diluted 10x in H_2O.

2.9.2 qRT-PCR analysis

qRT-PCR analysis was performed using either TaqMan or SYBR chemistry. For TaqMan probes, a reaction mix of 1 μl primer/probe mix (Invitrogen), 10 μl TaqMan Fast Advanced Master Mix (#4444557, Applied Biosystems), and 4 μl H_2O was added to each well of a 96 well reaction plate (#4346906, Applied Biosystems). For SYBR primers, a reaction mix of 0.5 μl 10 μM forward and reverse primers, 12.5 μl Fast SYBR Green Master Mix (#4385612, Applied Biosystems), and 7 μl H_2O was added to each well. 5 μl diluted cDNA was added to each reaction mix, and the plate was sealed with a MicroAmp Adhesive Film (#4311971, Applied Biosystems). The plate was briefly centrifuged, then loaded onto a QuantStudio 3 Real-time PCR machine (Applied Biosystems). TaqMan and SYBR reactions were run using the following cycling conditions: 95°C 20 s; 40 cycles of (95°C 1 s; 60°C 20 s). Ct values were analysed using the ΔΔCt method (Livak *et al.*, 2001), as detailed below, normalising to the housekeeping gene *GAPDH*. Primers and TaqMan probes are detailed in Table 2.6. Statistical analyses were performed using two-sided student's t-test.

$$\Delta Ct = Ct(Gene) - Ct(GAPDH)$$

$$\Delta\Delta Ct = \Delta Ct(sample) - \Delta Ct(reference)$$

$$\text{Fold difference} = 2^{-\Delta\Delta Ct}$$

Table 2.6: Primers and TaqMan probes used for gene expression qRT-PCR analysis

Target	Species	Primer sequence (5'-3') / Catalogue #	Chemistry	Note
GAPDH	Human	Hs03929097_g1	TaqMan	
RUNX1	Human	Hs00231079_m1	TaqMan	
CASP9	Human	Hs00609647_m1	TaqMan	
BCL2	Human	Hs00608023_m1	TaqMan	
MYC	Human	Hs0015348_m1	TaqMan	
MYB	Human	Hs00920556_m1	TaqMan	
BCL2 (intronic)	Human	F 5'-CGATAACGCCTGCCATCTAA-3' R 5'-CCACCACATCCTACTGGATTAC-3'	SYBR	
MLL-AF4	Human	F 5'-AGGTCCAGAGCAGAGCAAAC-3' R 5'-CGGCCATGAATGGGTCATTTC-3'	SYBR	
Ikzf1	Mouse	Mm01187880_m1	TaqMan	Used by L. Greder
Runx1	Mouse	Mm01213404_m1	TaqMan	Used by L. Greder

2.10 RNA-seq methods

2.10.1 Poly-A RNA-seq library preparation

For standard poly-A RNA-seq (as used in chapter 5), total RNA was extracted from 5 x 10^5 MLL-AF9 transformed GMPs (section 2.4.3) using an RNeasy Mini kit (74104, Qiagen), and RNA quality assessed by TapeStation. Poly-A RNA was isolated using the NEBNext Poly(A) mRNA magnetic isolation module (E7490, NEB). 1 μg RNA was incubated with oligo dT beads d(T)25 in 1X RNA binding buffer at 65°C 5 min then cooled to 4°C. Beads were washed 2x in wash buffer (E7490, NEB), resuspended in 50 μl Tris buffer, then incubated at 80°C 2 min to elute RNA. RNA was rebound to beads by addition of 1X RNA binding buffer and incubation at room temperature for 5 min. Libraries were then prepared using the NEBNext Ultra II Directional RNA library prep kit (E7760, NEB). Library quality was assessed using a DNA 1000 high sensitivity Screen Tape (Agilent 2200 TapeStation system) and quantified by qRT-PCR (KAPA library quantification kit KK4828). Samples were 150 bp paired-end sequenced by Novogene on a NovaSeq 6000. Tapestation traces, read counts and PCR duplication levels are displayed in Appendix A.7 and A.8.

2.10.2 Nascent RNA-seq library preparation

For nascent RNA-seq experiments (as used in chapter 4), 1 x 10^8 SEM or RS4;11 cells following siRNA treatment were treated with 500 μM 4-thiouridine (4-SU) for 1 hour at 37°C. Cells were lysed with TRIzol (#15596026, Life Technologies), and total RNA was extracted by chloroform and ethanol precipitation. DNA was degraded by DNase I treatment with the TURBO DNA-free kit (#AM1907). To biotinylate 4-SU incorporated RNA, up to 100 μg RNA was incubated with 200 μg/ml EZ-link HPDP-Biotin (#21341, Pierce) in biotinylation buffer (10 mM Tris-HCl pH 7.4, 1 mM EDTA) for 90 min at room temperature, followed by chloroform extraction. Biotinylated RNA was isolated by magnetic streptavidin bead pulldown. RNA was first incubated at 65°C 10 min, then incubated with an equal volume of magnetic streptavidin beads (μMACS Streptavidin Kit, #130-074-101, Miltenyi). RNA and beads were passed through a μMACS columns on magnet, then washed 3x with 900 μl 65°C wash buffer (100 mM Tris-HCl pH 7.5, 10 mM EDTA, 1 M NaCl, 0.1% Tween-20) and 3x room temperature wash buffer. RNA was eluted from columns by two rounds of 100 μl 100 mM DTT application. RNA was purified using a RNeasy MinElute kit (#74204, Qiagen). Libraries were then prepared using the NEBNext Ultra II Directional RNA library prep kit (E7760, NEB). Library quality was assessed using a DNA 1000 high sensitivity Screen Tape (Agilent 2200 TapeStation system) and quantified by qRT-PCR (KAPA library quantification kit KK4828), then 150 bp paired-end sequenced using a NextSeq 500 (Illumina). Read counts and PCR duplication levels are displayed in Appendix A.4.

2.10.3 Library preparation from embryonic populations

RNA-seq libraries from embryonic populations were prepared and sequenced by the WTCHG (Oxford). For each haematopoietic population from E8.5-E10.5 mouse embryos (see section 2.2.2), 100 cells were sorted directly into 3.16 μl lysis mix (Clontech), then used for library preparation by the WTCHG using the SMARTer kit (Clontech, #634935) for ultra-low RNA preparation. Library quality was assessed

using a DNA 1000 high sensitivity Screen Tape (Agilent 2200 TapeStation system) and quantified by qRT-PCR (KAPA library quantification kit KK4828). Libraries were then paired-end sequenced using an Illumina Hiseq2000. Read counts and PCR duplication levels are displayed in Appendix A.3.

2.11 Chromatin immunoprecipitation (ChIP) methods

2.11.1 ChIP

1 x $10^{7\text{-}8}$ cells were washed with PBS, then fixed. For assaying chromatin modifications, cells were fixed by incubation with 1 ml 1% formaldehyde at room temperature 10 min with rotation. For assaying TFs, cells were fixed by incubation with 1 ml 2 mM disuccinimidyl glutarate (DSG) at room temperature 30 min with rotation, followed by incubation in 1 ml 1% formaldehyde at room temperature 30 min with rotation. After fixation, cell pellets were washed with PBS, supernatant discarded, and snap frozen on dry ice. Fixed pellets were stored at -80°C.

For 1 x 10^8 cells, the following steps were split amongst 10 tubes, such that a maximum of 1 x 10^7 cells were sonicated at a time. Fixed pellets were resuspended in 120 μl SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris pH 8), with protease inhibitors. Cell pellets were sheared by passaging through a 27G needle, then incubated at 4°C 10 min. 120 μl lysed cell suspension was transferred to a Covaris microTUBE (#520045,Covaris) and sonicated for 5 min (for histone modifications) or 10 min (for TFs) using a Covaris ME220 (Woburn, MA, USA). Sonication parameters were set as follows: peak power = 75, duty factor = 25%, 500 cycles per burst. Sonicated samples were centrifuged at 20000 RCF 10 min 4°C, and supernatant collected. Samples were diluted 1:10 in ChIP dilution buffer (CDB) containing 0.01% SDS, 1.1% Triton X-100, 16.7 mM Tris-HCl pH 8.1, 167 mM NaCl, 1.2 mM EDTA. To pre-clear, samples were incubated with 5 μl of a 1:1 mixture of Protein A (#10001D, Invitrogen) and Protein G (#10004D, Invitrogen)

Dynabeads at 4°C 15 min with rotation. Samples were centrifuged at 1000 RCF 2 min, and pre-cleared supernatant collected. 1 ml aliquots were used for antibody incubation, and 50 μl aliquots for input material. 2 μl antibody (1:500 dilution) was added to 1 ml pre-cleared chromatin and incubated 4°C overnight with rotation. ChIP antibodies are detailed in Table 2.7.

Table 2.7: Antibodies used for ChIP experiments.

Target	Catalogue number	Company
RUNX1	ab23980	Abcam
MLL-N	A300-086A	Bethyl
AF4-C	ab31812	Abcam
MAZ	A301-652A	Bethyl
H3K27ac	C15410196	Diagenode
H3K4me1	pAb-194-050	Diagenode

All antibodies were raised in rabbit.

Ab:chromatin complexes were isolated using by incubating 1 ml pre-cleared chromatin + antibody with 15 μl of a 1:1 mixture of Protein A and Protein G Dynabeads (Invitrogen) at 4°C with rotation for at least 4 hours. Beads were washed 3x with RIPA buffer (50 mM HEPES-KOH pH 7.6, 500 mM LiCl, 1 mM EDTA, 1% IGEPAL, 0.7% Na deoxycholate), and washed 1x with 50 mM NaCl, 1 mM EDTA, and 10 mM Tris-HCl pH 8. Samples were eluted off beads by incubation with 1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8 at 65°C 30 min, centrifugation at 1000 RCF 2 min, and collection of supernatant. Samples were treated with 1 μl 10 mg/ml RNase A (#EN0531, Thermo Scientific) at 37°C 30 min, then 1 μl 20 mg/ml proteinase K with incubation at 65°C overnight to reverse DNA crosslinks. DNA was then purified using a QIAquick PCR purification kit (#28106, Qiagen) according to manufacturer's instructions.

2.11.2 ChIP qPCR

ChIP-qPCR analysis was performed with SYBR chemistry using primers detailed in Table 2.8. A reaction mix of 0.5 μl 10 μM forward and reverse primers, 12.5 μl

Fast SYBR Green Master Mix (#4385612, Applied Biosystems), and 7 μl H_2O was added to each well of a 96 well reaction plate (#4346906, Applied Biosystems). 5 μl DNA was added to each reaction mix, and the plate was sealed with a MicroAmp Adhesive Film (#4311971, Applied Biosystems). The plate was briefly centrifuged, then loaded onto a QuantStudio 3 Real-time PCR machine (Applied Biosystems). The plate was run using the following cycling conditions: 95°C 20 s; 40 cycles of (95°C 1 s; 60°C 20 s). ChIP signal was calculated as a percentage of input by the following calculation:

$$\Delta Ct = Ct(Input) - Ct(Sample)$$

$$\%Input = 2^{\Delta Ct} \times 5$$

2.11.3 ChIP-seq library preparation

1 x 10^8 cells were used for ChIP-seq experiments. For ChIP-seq on siRNA treated samples, reference normalisation was performed (Orlando *et al.*, 2014) by spiking fixed *Drosophila melanogaster* S2 cells prior to sonication in a 1:5 ratio. DNA libraries were generated using Ultra II DNA library preparation kit for Illumina (E7645, NEB) and 75 bp paired-end sequenced using a NextSeq 500 (Illumina). Read counts and PCR duplication levels are displayed in Appendix A.4.

Table 2.8: Primers used for ChIP qPCR analysis.

Target	Primer
Negative control (18205)	F - 5'-GGCTCCTGTAACCAACCACTACC-3'
	R - 5'-CCTCTGGGCTGGCTTCATTC-3'
+23 RUNX1 enhancer	F - 5'-TGCGAGAGCGAGAAAACCACAG-3'
	R - 5'-GCAGAAAGCAACAGCCAGAAACG-3'
CDK6 KMT2A-AFF1 peak	F - 5'-TCGAAGCGAAGTCCTCAACA-3'
	R - 5'-GCTTGGGCAGAGGCTATGTA-3'
BMF MLL-AF4/RUNX1 peak	F - 5'-AGAGCCAGAGTGCGTGAGAG-3'
BMF MLL-AF4/RUNX1 peak	F - 5'-CCACAGAGCAAACACAGG-3'

2.11.4 ChIPmentation library preparation

ChIPmentation was performed on low cell count samples in chapter 5, using a protocol adapted from Gustafsson *et al.*, 2019, and optimised by A. Smith (Milne lab, Crump *et al.*, 2022). 10 μl Protein A Dynabeads (#10001D, Invitrogen) were incubated with 1 μl antibody (Table 2.7) in 200 μl Binding Buffer (0.5% BSA in PBS with protease inhibitors) at 4°C 4 hours with rotation (referred to as Ab:beads here on). 7.5 x 10^5 MLL-AF9 transformed GMPs (section 2.4.3) were fixed under histone modification or TF specific conditions as described in section 2.11.1. Fixed cells were lysed in 120 μl lysis buffer (0.5% SDS, 50 mM Tris-HCl pH 8.0, 10 mM EDTA), then transferred to a Covaris microTUBE (#520045,Covaris) and sonicated as in section 2.11.1. 6 μl was collected as input chromatin. Additional lysis buffer was added to sonicated samples for a total volume of 240 μl per antibody (240 μl for TFs, 720 μl for histone modifications). Triton-X 100 was added for a final concentration of 1%, and samples were incubated 10 min room temperature to neutralise SDS. Samples were pre-cleared by incubation with 5 μl Protein A Dynabeads (Invitrogen) at 4°C 30 min with rotation. Ab:beads incubated as above were washed with 0.5% FCS/PBS. 240 μl pre-cleared chromatin was added to Ab:beads, then incubated at 4°C overnight with rotation. Of 7.5 x 10^5 cells prepared for histones, three antibodies were used (H3K4me3 ChIPmentation not shown), such that each histone modification ChIPmentation represents 2.5 x 10^5 cells.

Samples were washed 3x with RIPA buffer, 1x with 1 mM EDTA and 10 mM Tris-HCl pH 8.1, and 1x. with 10 mM Tris-HCl pH 8.1. Chromatin was tagmented by incubating beads in 29 μl Tagmentation Buffer (10 mM Tris-HCl pH 8.1, 5 mM $MgCl_2$, 10% dimethylformamide (DMF)) and 1 μl Tn5 transposase (#20034197, Illumina) at 37°C for 10 min. After incubation samples were diluted with 150 μl RIPA buffer, then washed with 150 μl 10 mM Tris-HCl pH 8.0. Beads were then resuspended in 22.5 μl H_2O. 22.5 μl tagmented chromatin DNA was indexed by adding 25 μl NEBNext Ultra II Q5 Master Mix (#M0544S, NEB) + 1.25 μl 5 μM Universal adapter + 1.25 μl 5 μM custom index primer, then PCR amplified with

the following cycling conditions: 72°C 5 min, 95°C 5 min, 11x cycles (98°C 10 s, 63°C 30 s, 72°C 3 min). Library clean-up was performed with Agencourt AMPure XP beads at a 1:1 ratio. Library quality was assessed using a DNA 1000 high sensitivity Screen Tape (Agilent 2200 TapeStation system) and quantified by qRT-PCR (KAPA library quantification kit KK4828). Samples were 75 bp paired-end sequenced by paired-end sequencing on a NextSeq 500 (Illumina). Tapestation traces, read counts and PCR duplication levels are displayed in Appendix A.7 and A.8.

2.12 ATAC-seq library preparation

ATAC-seq libraries were generated from E8.5 - E10.5 mouse embryos by L. Greder (de Bruijn lab) and M. Suciu (Hughes lab), and from MLL-AF9 transformed GMPs myself, using a protocol adapted from (Buenrostro *et al.*, 2015). For mouse haematopoietic populations (section 2.2.2), 1500 cells were sorted into 100% FCS (Gibco). For MLL-AF9 transformed GMPs, 1 x 10^5 cells were harvested from MethoCult cultures (section 2.4.3). Cells were washed in PBS, then resuspended in lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL). Cells were centrifuged 500 RCF 10 min 4°C, then resuspended with 2.5µL Tn5 transposase (#20034197, Illumina) in 25 μl 2xTD buffer (#20034197, Illumina) and 22.7 μl H_2O and incubated at 37°C 30 min. Tagmentation reaction for embryonic populations was quenched with 1.1 μl 500 mM EDTA, and supplemented with 10 μl 50 mM $MgCl_2$. Tagmentation reaction for MLL-AF9 GMPs was stopped by clean-up with a MinElute PCR Purification Kit (#28004, Qiagen), according to manufacturer's instructions, with elution in 20 μl 10 mM Tris-HCl pH 8.1. 20 μl tagmented DNA was amplified and indexed by adding 25 μl NEBNext Ultra II Q5 Master Mix (#M0544S, NEB) + 2.5 μl 25 μM Universal adapter + 2.5 μl 25 μM custom index primer, then PCR amplified with the following cycling conditions: 72°C 5 min, 95°C 30 s, 8x cycles (98°C 10 s, 63°C 30 s, 72°C 1 min). PCR clean-up was performed using a MinElute PCR Purification Kit (#28004, Qiagen), and eluted in 20 μl H_2O. Library quality was assessed using a DNA 1000

high sensitivity Screen Tape (Agilent 2200 TapeStation system) and quantified by qRT-PCR (KAPA library quantification kit KK4828). Libraries were 75 bp paired-end sequenced on a NextSeq 500 (Illumina), or were 150 bp paired-end sequenced by Novogene on a NovaSeq 6000.

2.13 RNA+ATAC multiome library preparation

Cells for RNA+ATAC multiome library preparation were acquired by S. Rice and N. Elliott (Roy lab) and nuclei isolated by A. Smith (Milne lab). Patient and donor samples are detailed in section 4.8, p.136. Patient and donor samples were prepared for sorting as previously described (O'Byrne *et al.*, 2019). Samples were stained with an a fetal bone marrow (FBM) and fetal liver (FL) antibody panel (Table 2.9), and ALL blast antibody panel (Table 2.10), and sorted using by FACS on a BD FACSAria Fusion. Cells were sorted for the following phenotypes: FL & FBM CD34+ ($Lin^-CD34^+CD38^-$ and $Lin^-CD34^+CD38^+$); FL & FBM CD34- ($CD34^-$); ALL blasts ($CD19^+$). Samples were pooled in pairs, such that 1x male and 1x female sample were combined (Table 4.1, p.137).

Nuclei were isolated from pooled samples according to the 10X Genomics nuclei isolation protocol (CG000365), using the Low Cell Input Nuclei Isolation method. Cells were centrifuged at 300 RCF 5 min 4°C and resuspended in 0.04% BSA/PBS. 4 x 10^4 cells were taken and diluted to 50 μl in 0.04% BSA/PBS, then centrifuged at 300 RCF 5 min 4°C. The pellet was resuspended in lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 0.1% IGEPAL, 0.01% Digitonin (#BN2006, Life technologies), 1% BSA, 1 mM DTT, 1 U/μl RNase inhibitor (Sigma)), and incubated for 3 min at 4°C. To stop lysis, 50 μl wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 1% BSA, 0.1% Tween-20, 1 mM DTT, 1 U/μl RNase inhibitor (Sigma)) was added to samples, then centrifuged 500 RCF 5 min 4°C. Pellet was resuspended in 45 μl diluted nuclei buffer (1X Nuclei Buffer (10x Genomics) with 1 mM DTT and 1 U/μl RNase inhibitor (Sigma)),

Table 2.9: Antibody panel for sorting fetal bone marrow and liver donor cells.

Target	Conjugate	Dilution	Catalogue number	Company
CD10	FITC	1/20	11-0106-42	Life Tech
CD90	BV421	1/20	328122	Biolegend
CD34	Pecy7	1/50	25-0349-42	Life Tech
CD45RA	APCef780	1/50	47-0458-42	Life Tech
CD127	PE	1/50	12-1278-42	Life Tech
CD123	BV650	1/50	563405	BD Bioscience
CD2	PerCP-Cy5.5	1/50	300216	Biolegend
CD3	PerCP-Cy5.5	1/50	317336	Biolegend
CD14	PerCP-Cy5.5	1/50	301824	Biolegend
CD16	PerCP-Cy5.5	1/50	302028	Biolegend
CD56	PerCP-Cy5.5	1/50	318322	Biolegend
CD235	PerCP-Cy5.5	1/100	306614	Biolegend
CD19	APC	1/100	302212	Biolegend
CD38	af700	1/100	56-0389-42	Life Tech
Viability	ef506	1/200	65-0866-18	Life Tech

then centrifuged 500 RCF 5 min 4°C. Pellet was resuspended in 7 μl diluted nuclei buffer. 2 μl nuclei suspension was mixed with 8 μl diluted nuclei buffer and 10 μl Trypan Blue and counted with a haemocytometer to ensure sufficient nuclei quality and quantity. Multiome libraries were generated by an in-house facility (performed by Neil Ashley) following the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression protocol (CG000338, 10X Genomics). Samples were 150 bp paired-end sequenced by Novogene on a NovaSeq 6000.

2.14 Computational analysis

2.14.1 RNA-seq analysis

FASTQ files were quality checked using FastQC (v0.11.4) and trimmed using trim_galore (v0.4.1) (Martin, 2011). Paired-end reads were mapped to mm9 (chapter 3), hg19 (chapter 4), or mm10 (chapter 5) reference genomes using STAR (v2.4.2) (Dobin *et al.*, 2013). Reads mapping to ENCODE blacklisted regions (Amemiya *et al.*, 2019) were discarded using BEDtools (v2.17.0) (Quinlan

Table 2.10: Antibody panel for sorting ALL blast samples.

Target	Conjugate	Dilution	Catalogue number	Company
CD10	FITC	1/20	11-0106-42	Life Tech
CD34	PEcy7	1/50	25-0349-42	Life Tech
CD2	PerCP-Cy5.5	1/50	300216	Biolegend
CD3	PerCP-Cy5.5	1/50	317336	Biolegend
CD14	PerCP-Cy5.5	1/50	301824	Biolegend
CD16	PerCP-Cy5.5	1/50	302028	Biolegend
CD56	PerCP-Cy5.5	1/50	318322	Biolegend
CD235a	PerCP-Cy5.5	1/100	306614	Biolegend
CD19	APC	1/100	302212	Biolegend
CD20	ef450	1/100	48-0209-42	Life Tech
CD38	BV605	1/100	303532	Biolegend
CD45	AF700	1/100	56-9459-42	Life Tech
CD133	PE	1/100	130-113-108	Miltenyi
Viability	ef506	1/200	65-0866-18	Life Tech

et al., 2010). PCR duplicates were removed using Picard-tools MarkDuplicates (v1.83). Gene expression was quantified by counting mapped reads over exons using subread featureCounts (v1.6.2) (Liao *et al.*, 2014) and lowly expressed genes were discarded, to retain genes with > 3 counts per million (CPM) in at least 3 samples. Statistical analyses were performed in R (R Core Team, 2021) using the edgeR package (Robinson *et al.*, 2010), with benjamini and hochberg multiple hypothesis correction. Statistical comparisons involved simple pairwise comparisons between two samples, or "ANOVA-like" analysis (Robinson *et al.*, 2010) to compare across all samples. Where blocking factors were used to remove the effect of unwanted variation (e.g. experimental batches), experimental design was modelled in edgeR as: model.matrix(∼sampleGroup + blockingFactor). Differentially expressed genes (DEGs) were defined as FDR ≤ 0.05. Gene expression correlations were calculated using Spearman's rank correlation in R.

2.14.2 ChIP-seq and ATAC-seq analysis

FASTQ files were either processed using the NGseqBasic pipeline (Telenius *et al.*, 2018) as in chapters 3 and 4, or with a custom pipeline as in chapter 5. FASTQ files were mapped to mm9 (chapter 3), hg19 (chapter 4), or mm10 (chapter 5) reference genomes.

For NGseqBasic pipeline, FASTQ files were quality checked using FastQC (v0.11.4) and mapped using Bowtie (v1.0.0) (Langmead, 2010). Unmapped reads were trimmed with trim_galore (v0.3.1) (Martin, 2011), and short unmapped reads were combined using flash (v1.2.8) (Magoč *et al.*, 2011) and subsequently remapped. PCR duplicates were removed using SAMtools rmdup (v0.1.19) (Li *et al.*, 2009), and reads mapping to Duke blacklisted regions (UCSC) were discarded using BEDtools (v2.17.0) (Quinlan *et al.*, 2010). For custom pipeline, FASTQ files were quality checked using FastQC (v0.11.9) and mapped using bowtie2 (v2.4.5) (Langmead *et al.*, 2012). Reads mapping to ENCODE blacklisted regions (Amemiya *et al.*, 2019) were discarded using BEDtools (v2.30.0) (Quinlan *et al.*, 2010). PCR duplicates were removed using deepTools alignmentSieve (v3.4.3) (Ramírez *et al.*, 2016). For reference-normalised ChIP-seq, FASTQ reads were additionally mapped to the dm6 reference genome. hg19 read counts were adjusted based on the ratio of drosophila to human reads in input and IP control/treatment samples.

Tag (read) directories and bigWigs were made using HOMER (v4.7) (Heinz *et al.*, 2010), with counts normalised to tags per 1 x 10^7 tags, and visualised on the UCSC genome browser (Kent *et al.*, 2002). For the purposes of visualisation, ATAC-seq and ChIPmentation replicates were merged prior to bigWig generation and normalisation. ChIP-seq peak calling was performed using HOMER findPeaks with input tracks for background correction. ATAC-seq peak calling was performed using MACS2 (v2.0.1/v2.2.7.1) (Zhang *et al.*, 2008) with parameters -BAMPE and -q 0.05. Note that MLL-AF4 peak calls were defined as MLL-N ChIP-seq peaks overlapped with AF4-C ChIP-seq peaks. Where peak call overlaps were

performed, overlaps are defined as regions with at least 1 bp overlay. Called peaks were annotated to the nearest promoter using either HOMER annotatePeaks.pl (v4.7), or a custom approach (see section 2.15.4). For ChIP-seq data, the nearest annotated gene is referred to as a "bound gene". Heatmaps and scatter plots profiling ChIP-seq and ATAC-seq profiles were made using deepTools (v3.4.3).

For statistical quantification of ATAC-seq data, robust peaks were first identified. Robust peaks are defined as elements where at least 2 samples are overlapping, such that peak calls in just one replicate are discarded. A union set of peaks was generated using the R package DiffBind dba.count, which normalises peak widths as a 400 bp window centred over the summit of ATAC-seq signal. Read counts were summarised across the union peak set, and lowly accessible peaks were discarded, to retain peaks with > 3 CPM in at least 3 samples. Statistical analyses were performed in R using the edgeR package (Robinson *et al.*, 2010), with benjamini and hochberg multiple hypothesis correction. Statistical comparisons involved simple pairwise comparisons between two samples, or "ANOVA-like" analysis (Robinson *et al.*, 2010) to compare across all samples. Differentially accessible elements (DAEs) were defined as FDR ≤ 0.05. ATAC-seq and ChIP-seq correlations were calculated using Spearman's rank correlation in R.

2.14.3 Dimensionality reduction methods

RNA-seq and ATAC-seq datasets throughout are dimensionally reduced using principal component analysis (PCA) and UMAP was performed using R (R Core Team, 2021). RNA-seq and ATAC-seq input was first variance stabilised using DESeq2 vst (Love *et al.*, 2014). PCA loading scores were also extracted, to represent genes that confer the greatest influence on principal components in PCA plots.

2.14.4 Single-cell sequencing analysis

Published single cell RNA-seq (scRNA-seq) of mouse and human EHT (Table 2.11, Zeng *et al.*, 2019; Zhu *et al.*, 2020) was reanalysed in chapter 3, and

RNA+ATAC multiome data (generated as in section 2.13) was analysed in chapter 4. EHT scRNA-seq was mapped to mm9, and RNA+ATAC multiome was mapped to hg19 reference genomes. For STRT-seq data, pre-processed expression tables were downloaded. For 10X genomics data, scRNA-seq was mapped using 10x Genomics Cell Ranger (v6.0.1) (Zheng *et al.*, 2017), and RNA+ATAC multiome was mapped using 10x Genomics Cell Ranger ARC (v2.0.0). scRNA-seq expression was pre-processed using the R package Seurat (v4.1.0) (Hao *et al.*, 2021), and scATAC-seq accessibility was pre-processed using ArchR (v1.0.1) (Granja *et al.*, 2021).

For RNA+ATAC multiome, samples were pooled in male-female pairs (section 2.13). To demultiplex cells, the souporcell pipeline (Heaton *et al.*, 2020) was used to classify cells based on genotype, and ambiguously assigned classes were discarded. Gender identities were assigned to souporcell classifications by examining the number of *XIST* transcripts and % of reads mapped to the Y chromosome. Male samples were assigned as high Y chromosome mapping, and female as high *XIST* expression (Appendix A.5). Two samples pooled were both male (iALL3 and chALL, see Table 4.1). For these samples, souporcell classifications were assigned by cross-referencing variant calls by souporcell with ATAC-seq and ChIP-seq datasets generated within the Milne lab (data not shown).

EHT scRNA-seq cells were filtered as in their respective publications (Zeng *et al.*, 2019 and Zhu *et al.*, 2020). RNA+ATAC multiome cells were filtered based on expression data by outlier detection using scater (v1.22.0) (McCarthy *et al.*, 2017), and additionally retaining cells with the following parameters: read counts > 300 & < 50000; detected features > 700 & < 7000, % mitochondrial reads $< 20\%$ (Appendix A.5). Cells were also filtered using the scATAC-seq data, to retain cells with a TSS enrichment score greater than 4 and greater than 1000 fragments. For all datasets, scRNA-seq expression was normalised and scaled using the SCTransform method in Seurat (Hafemeister *et al.*, 2019), with variable regression on % mitochondria, total read counts, and total detected features.

scRNA-seq data was dimensionally reduced by PCA, followed by UMAP analysis. scATAC-seq data was dimensionally reduced by iterative latent semantic indexing (LSI, as used in Granja *et al.*, 2019; Satpathy *et al.*, 2019), followed by UMAP analysis. For FBM and FL data, reciprocal PCA (RPCA) analysis was performed with Seurat to batch correct for donor identity. To integrate scRNA-seq and scATAC-seq datasets together, weighted nearest neighbour analysis (WNN, Hao *et al.*, 2021) was performed using Seurat FindMultiModalNeighbors (v4.1.0), with PCA and ILSI graphs as input. This WNN graph was used for WNN-UMAP analysis. Multimodal clusters were identified with Seurat, using WNN graphs as input. To label multimodal clusters, a scRNA-seq reference dataset was assembled by integrating and RPCA batch correcting published FBM and FL datasets (Table 2.11, Jardine *et al.*, 2021; Popescu *et al.*, 2019; Roy *et al.*, 2021). Labels were transferred from this reference dataset to the multiome scRNA-seq dataset using Seurat commands FindTransferAnchors and TransferData. B-cell populations were manually annotated using gene expression as used in Jardine *et al.*, 2021 and Suo *et al.*, 2022 (Appendix A.6).

2.14.5 Pseudotime analysis

EHT scRNA-seq data was dimensionally reduced using either ForceAtlas2 (FA2) force directed graph (Jacomy *et al.*, 2014), or UMAP analysis. RNA+ATAC multiome data was dimensionally reduced using WNN UMAP analysis. The R package Slingshot (Street *et al.*, 2018) was used to calculate pseudotime trajectories from UMAP and FA2 coordinates. The R package TradeSeq (Van den Berge *et al.*, 2020) with parameter k.knots = 7, was used to fit a model $\log_e$ normalised expression as a function of pseudotime. Differential analysis was performed using TradeSeq associationTest to calculate genes that change in expression as pseudotime increases, with strength of significance represented by the wald statistic. Significantly associated genes were clustered by grouping gene expression data over 100 bins, and clustering with k-means in R. AUCell algorithm (Aibar *et al.*, 2017) was used to calculate single-cell enrichment scores for TradeSeq derived gene expression clusters.

scRNA-seq gene expression correlation was performed by first reducing noise using a moving average of gene expression over pseudotime (with a 20-cell averaging window), then Spearman's rank correlation using R.

2.14.6 Pathway enrichment

For DEGs, enriched GO terms and reactome pathways were determined using PANTHER (v15) (Thomas *et al.*, 2003). For DAEs, enriched GO terms were determined using Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean *et al.*, 2010).

2.15 GRN construction and analysis

2.15.1 TF binding motif analysis

All motif analyses were performed using the MEME Suite (v5.4.1) (Bailey *et al.*, 2015), and input ATAC-seq peaks were reduced to a 200 bp window. TF binding motifs were downloaded from the HOCOMOCO database (Kulakovskiy *et al.*, 2018), which included full and core databases. The core database included only higher confidence motif positional weight matrices (PWMs). For FIMO analysis used in GRN creation the core HOCOMOCO database was used, and was filtered for TF motifs associated with DEGs to restrict the size of the database. To perform TF binding motif enrichment using MEME AME (McLeay *et al.*, 2010), motifs from the full HOCOMOCO database were used, using all accessible elements as the control peak set. TF binding motifs were detected within individual ATAC-seq peaks using MEME FIMO (Grant *et al.*, 2011).

2.15.2 EHT GRN methodology

A GRN model of mouse EHT was established by integrating RNA-seq and ATAC-seq datasets. See Appendix A.9 for a flowchart overview of the steps. Enhancer-promoter (E-P) associations were predicted by annotating DAEs to the nearest DEG

promoter, using HOMER (v4.7) annotatePeaks.pl (Heinz *et al.*, 2010), such that differential ATAC-seq peaks are linked with differential gene promoters. Pearson's correlation was calculated between ATAC-seq $\log_2$ CPM of the DAE and RNA-seq $\log_2$ CPM of the linked DEG, and E-P links were filtered for correlation greater than 0.4 or lower than -0.4. This correlation filter enriches for true enhancers or silencers, and increases the likelihood that the accessible element regulates the linked gene.

To predict upstream regulators, TF binding motifs underlying each ATAC-seq peak were predicted using MEME FIMO (Grant *et al.*, 2011) and the core HOCOMOCO database (Kulakovskiy *et al.*, 2018), as described in section 2.15.1. Only TF motifs corresponding with differentially expressed TFs were included. Presence of a TF motif implies a TF:target interaction via the E-P link. Predicted TF:target interactions were filtered based on RNA-seq $\log_2$ CPM, retaining interactions with greater than 0.4 or lower than -0.4 correlation. Predicted E-P associations and motif analyses were integrated into a network format using the R package iGraph (v1.2.11). PCA gene loadings coordinates (section 2.14.3) were annotated to each GRN node and used to layout the network for visualisation purposes, such that the x axis describes PC1 (EHT progression), and the y axis describes PC2 (transient expression).

2.15.3 MLL-AF4 leukaemia GRN methodology

GRN models describing MLL-AF4 leukaemia were generated using ChIP-seq and nascent RNA-seq following gene siRNA KD. See Fig. 4.3A, p.110 for an illustration of the methodology. Datasets sourced from published work are detailed in Table 2.11. GRNs were created to describe the regulatory impact of a specific factor: MLL-AF4 or RUNX1. This method involved four steps: 1). The regulatory scope of MLL-AF4 or RUNX1 was estimated using nascent RNA-seq following 96 hours' *MLL-AF4* or *RUNX1* KD. 2). It was assumed that individual TFs have both direct and indirect regulatory effects. Direct interactions were established using MLL-AF4 and RUNX1 ChIP-seq peak calls, which were annotated to the

nearest expressed gene promoter using HOMER annotatePeaks.pl (v4.7). ChIP-seq peaks were additionally filtered for regions overlapping H3K27ac and H3K4me1 to enrich for enhancers. Genes annotated in this way were referred to as MLL-AF4 or RUNX1 "bound" genes. 3). To assign indirect interactions, a more general predicted TF interaction network was incorporated. This supplementary network was sourced from a FANTOM consortium EdgeExpressDB (Severin *et al.*, 2009; The FANTOM Consortium and the Riken Omics Science Center, 2009), which contains predicted TF interactions at gene promoters and was constructed using the THP-1 cell line. The R package GeneNetworkBuilder (v1.26.1) was used to assign predicted TF interactions from the FANTOM TF network to MLL-AF4/RUNX1 bound TFs, such that indirect regulation can be modelled (see Fig. 4.3A, p.110 for a schematic representation). 4). Nodes of the network were then annotated using ALL and AML patient RNA-seq (Agraz-Doblas *et al.*, 2019; O'Byrne *et al.*, 2019; The Cancer Genome Atlas Research Network, 2013), CRISPR screens (Behan *et al.*, 2019; Tzelepis *et al.*, 2016), and measures of centrality (section 2.15.5)

Patient expression data (Table 2.11, Agraz-Doblas *et al.*, 2019; O'Byrne *et al.*, 2019; The Cancer Genome Atlas Research Network, 2013) was used to create sub-networks from the MLL-AF4 GRN model. DEGs from the MLL-AF4 siRNA KD nascent RNA-seq data were filtered for genes active in each RNA-seq patient sample. The expression threshold to define "active" genes was determined for each dataset separately, and was calculated as the mean expression value across all genes and samples (See Fig. 4.5B, p.114 for expression histograms and thresholds). Note that median expression was not viable due to low sequencing depth in the FBM data. Node activity was further summarised as a binary matrix (individual samples x gene) and k-means clustering was performed to identify patient GRN modules.

2.15.4 MLL-AF9 GMP leukaemogenesis GRN methodology

A GRN model of MLL-AF9 GMP leukaemogenesis was established by integration RNA-seq, ATAC-seq, and ChIPmentation datasets. See Appendix A.10 for a

flowchart overview of the steps. First, ATAC-seq peaks were filtered for those overlapping H3K27ac ChIPmentation peaks, to mark enhancer elements. E-P associations were predicted using a custom iterative peak calling approach (see Fig. 5.5A for an illustration). This involves 5 steps: 1). DAEs were linked to all expressed promoters in a 50 kb window. 2). Any enhancers that are sufficiently annotated (at least 3 promoter annotations) are "resolved." 3). The current window was expanded +10 kb either side, and unresolved elements re-annotated. 4). Steps 2 and 4 are repeated until all peaks are resolved, or the annotation window reaches 3 Mb. 5). Remaining unannotated peaks were assigned to the nearest expressed promoter. Pearson's correlation was calculated between ATAC-seq $\log_2$ CPM of the DAE and RNA-seq $\log_2$ CPM of linked DEGs, and E-P links were filtered for correlation greater than 0.4 or lower than -0.4. E-P links were further filtered using TAD domains from the HPC7 cell line (Pinto do O *et al.*, 1998). E-P links were retained where both the enhancer element and the associated promoter are within the same TAD.

To predict upstream regulators, TF binding motifs underlying each ATAC-seq peak were predicted using MEME FIMO (Grant *et al.*, 2011) and the core HOCOMOCO database (Kulakovskiy *et al.*, 2018), as described in section 2.15.1. Only TF motifs corresponding with differentially expressed TFs were included. Presence of a TF motif implies a TF:target interaction via the E-P link. MLL-N ChIPmentation data (representing MLL-AF9) was also integrated, to predict MLL-AF9 regulatory interactions. Predicted TF:target interactions were filtered based on RNA-seq $\log_2$ CPM, retaining interactions with greater than 0.4 or lower than -0.4 correlation. Predicted E-P associations and motif analyses were integrated into a network format using the R package iGraph (v1.2.11).

2.15.5 Centrality calculation

Network degree and stress centralities were calculated using the R packages iGraph (v1.2.11) and sna (v2.4). Stress fold change (stress FC) was calculated in three steps: 1). in silico deletion of one GRN node at a time. 2). Recalculation

of stress centrality for all other genes. 3). Calculation of fold change in stress centralities after node deletion. Mean stress FC was determined by averaging stress FC values across all genes other than the in silico deleted node.

2.15.6 TF Co-interaction analysis

TF co-interaction is defined here as a pair of TFs that are predicted to interact at the same accessible elements in a network model, at a higher frequency than would be expected if TF interactions were random. For every pair of TFs (TFa and TFb), a contingency table summarising accessible elements targeted by TFa or TFb were extracted. A Fisher's exact test was performed comparing TFa elements and TFb elements, using all GRN elements as a background set, and giving a P-value and odds ratio. Odds ratio describes the likelihood that two TFa and TFb will target the same element. P-values were multiple-hypothesis corrected using the benjamini and hochberg method to generate FDR values. TFs not significant for any co-interaction were discarded (FDR > 0.05). Odds ratios were then clustered by hierarchical clustering in R.

2.16 Statistics

Unless otherwise stated, statistical comparisons were performed using two-sided student's t-test. For all box plot visualisations, midline indicates median, and upper and lower hinges indicate 1st and 3rd quartiles, respectively. Whiskers extend to furthest values within 1.5 times the interquartile range.

2.17 Sourcing of patient samples

ChIP-seq data generated in two MLL-AF4 patient samples were used in chapter 4. MLL-AF4 patient sample #1 is described in Kerry *et al.*, 2017. MLL-AF4

patient sample #2 is a primary diagnostic bone marrow sample from a 6-year-old child with ALL, obtained from the Bloodwise Childhood Leukaemia Cell Bank, UK (REC: 16/SW/0219). Samples were anonymized at source, assigned a unique study number and linked.

2.18 Sourcing of published datasets

Published datasets used in thesis are detailed in Table 2.11.

Table 2.11: Published datasets used in this thesis, detailing publication and GEO accession numbers.

Experiment	Cell line/tissue	Antibody	Treatment	Accession	Publication
Tiled capture-C	mESC diff.			GSE184490	(Owens *et al.*, 2022)
ChIP-seq	416B	H3K27ac		GSE69776	(Schütte *et al.*, 2016)
ChIP-seq	416B	Runx1		GSE69776	(Schütte *et al.*, 2016)
ChIP-seq	SEM	MLL-N		GSE74812	(Kerry *et al.*, 2017)
ChIP-seq	SEM	AF4-C		GSE74812	(Kerry *et al.*, 2017)
ChIP-seq	SEM	RUNX1		GSE42075	(Wilkinson *et al.*, 2013)
ChIP-seq	SEM	Input		GSE42075	(Wilkinson *et al.*, 2013)
ChIP-seq	SEM	ELF1		GSE117865	(Godfrey *et al.*, 2019)
ChIP-seq	SEM	H3K79me3		GSE117865	(Godfrey *et al.*, 2019)
ChIP-seq	SEM	BRD4		GSE83671	(Kerry *et al.*, 2017)
ChIP-seq	SEM	H3K27ac		GSE74812	(Kerry *et al.*, 2017)
ChIP-seq	SEM	MYB		GSE117865	(Godfrey *et al.*, 2019)
ChIP-seq	SEM	ERG		GSE117865	(Godfrey *et al.*, 2019)
ChIP-seq	Primograft	KMT2A-N		GSE83671	(Kerry *et al.*, 2017)
ChIP-seq	Primograft	AFF1-C		GSE83671	(Kerry *et al.*, 2017)
ATAC-seq	SEM			GSE74812	(Kerry *et al.*, 2017)
Nascent RNA-seq	SEM		MLL-AF4 KD	GSE85988	(Godfrey *et al.*, 2017)
Nascent RNA-seq	SEM		EPZ-5676	GSE83671	(Kerry *et al.*, 2017)
Nascent RNA-seq	SEM		IBET	GSE139437	(Crump *et al.*, 2021)
10x scRNA-seq	Mouse EHT			GSE135202	(Zhu *et al.*, 2020)
STRT-seq	Human EHT			GSE137117	(Zeng *et al.*, 2019)
10x scRNA-seq	FL			E-MTAB-7407	(Popescu *et al.*, 2019)
10x scRNA-seq	FBM			GSE166895	(Jardine *et al.*, 2021)
10x scRNA-seq	FL & FBM			GSE155259	(Roy *et al.*, 2021)
RNA-seq	MLLr ALL			PRJEB23605	(Agraz-Doblas *et al.*, 2019)
RNA-seq	AML			NA	(The Cancer Genome Atlas Research Network, 2013)
RNA-seq	FBM			GSE122982	(O'Byrne *et al.*, 2019)

3

A regulatory network model underlying the endothelial-to-haematopoietic transition

Contents

3.1 Introduction and aims 63
3.2 Profiling of EHT sub-populations reveals dynamic transcription and chromatin accessibility changes . . . 65
3.2.1 Different EHT expression signatures drive differentiation, morphological, and signalling processes 66
3.2.2 Transition-specific regulation 70
3.3 Generating, and characterising a GRN model of EHT 73
3.3.1 Central regulators and targets of the EHT GRN 75
3.4 *Runx1* regulators and target genes 75
3.4.1 Cis-regulatory landscape of the *Runx1* locus 76
3.4.2 Predicted regulators of *Runx1* expression 78
3.4.3 Predicted targets of Runx1 81
3.5 Runx1 cooperates with TFs to drive EHT processes . 83
3.6 Potential Runx1 driven FFLs in EHT 88
3.6.1 A Runx1 and Ikzf1 driven FFL regulates Notch pathway genes 90
3.7 Maf and Lef1 cooperation may drive cytoskeletal remodelling and EMT processes 94
3.8 Conclusions and discussion 98

3.1 Introduction and aims

EHT is an essential process for the initial formation of the haematopoietic system. It involves the specification of endothelial cells into haemogenic endothelium and differentiation into HSCs and haematopoietic progenitor cells (HPCs) (Bertrand *et al.*, 2010; Boisset *et al.*, 2010; Kissa *et al.*, 2010). EHT involves cell differentiation through several distinct intermediates, extensive morphological changes, and genome wide transcriptional changes. Exploring how the transcriptome is regulated during EHT is essential for understanding what drives this process. Additionally, the well-defined cell types and dynamic transcription offers this system as an ideal model to explore the activity of TFs, and how these regulators drive specific processes and cell-type transitions underlying EHT.

It is well established that Runx1 is a master regulator critical for EHT (Cai *et al.*, 2000; Chen *et al.*, 2009; North *et al.*, 1999). It is therefore important to understand the context of Runx1 within the EHT process, such as its key upstream regulators and downstream targets. TFs do not act in isolation, and commonly function together. For example, Runx1 is known to cooperate with many different TFs, such as PU.1 and Ets1 (Goetz *et al.*, 2000; Hu *et al.*, 2011; Kim *et al.*, 1999b), and can alter the DNA binding preference of other factors as with Fli1 (Lichtinger *et al.*, 2012). More recently, in T-cell development Runx1 has been found to cooperate with Gata3 to drive a silencer for *Spi1* (Hosokawa *et al.*, 2021), and to cooperate with Runx3 to additively regulate gene expression with high functional redundancy (Shin *et al.*, 2021). Exploring how Runx1 cooperates with other TFs will offer both greater understanding of EHT, as well as insight into which other TFs cooperate to drive gene expression.

To explore the regulation of EHT in this chapter, I have constructed a GRN model integrating RNA-seq and ATAC-seq data. A GRN model is an ideal method to interrogate this system, as it allows us to determine the key drivers of specific processes, as well as to explore the combinatorial logic of TFs functioning

cooperatively. Several studies have attempted to characterise the regulatory network underpinning EHT (Baron *et al.*, 2018; Bergiers *et al.*, 2018; Gao *et al.*, 2020; Goode *et al.*, 2016; Zeng *et al.*, 2019; Zhu *et al.*, 2020), however these studies have primarily focused on E10.5 and E11.5 mouse embryos and have not investigated E8.5 mouse embryos where HE signals are first received. As such, the initial transcriptional changes occurring in early-EHT intermediates have been poorly studied, and lack characterisation of the E8.5 pre-HE population. In this chapter I aimed to construct a GRN model incorporating both early (pre-HE) and late (pre-HSC) populations, covering the full range of EHT intermediates.

Aims

1. Establish a GRN describing early and late endothelial-to-haematopoietic transition.
2. Profile regulators of *Runx1* and downstream targets.
3. Investigate transcription factor cooperativity that drives haematopoietic processes.

By constructing a GRN describing both early and late EHT, and profiling the regulators of Runx1, this analysis will improve our understanding of how EHT is initiated in the early stages of HE specification. Investigating TF cooperativity, will allow us to explore a less understood aspect of gene regulation that may be important in driving key processes.

3.2 Profiling of EHT sub-populations reveals dynamic transcription and chromatin accessibility changes

To probe EHT in detail the de Bruijn lab uses a *Runx1* +23 enhancer GFP reporter (23GFP) mouse model (Bee *et al.*, 2010), with specific activity in HE and IACs, allowing us to dissect early stages of EHT. 23GFP activity marks HE (Fig. 3.1A), and can be used to label a pre-HE population in E8.5 dorsal aorta endothelium that precedes active *Runx1* transcription (Swiers *et al.*, 2013a). This early activation of the +23 enhancer implies early regulatory signals prior to the onset of endogenous *Runx1* expression and EHT. As such, by using the 23GFP mouse model, HE can be isolated from E9.5 embryos and pre-HE can be isolated from E8.5 embryos, with a key difference between these populations being the detection of endogenous *Runx1* transcripts (Fig. 3.1A, Swiers *et al.*, 2013a).

To investigate the regulatory logic driving EHT I used data generated previously in the de Bruijn lab, namely transcription profiling and chromatin accessibility of bulk populations from the PAS/AGM + VU of E8.5 to

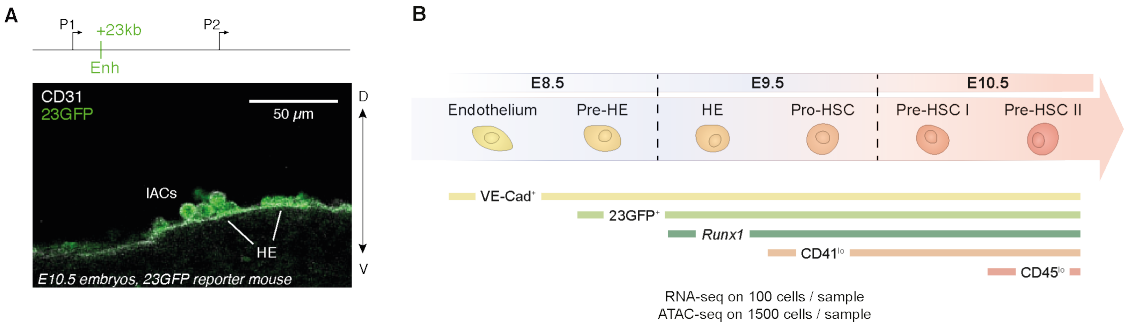


Figure 3.1: Overview of profiled EHT populations. (A) Top – Schematic of the *Runx1* locus, with the +23 enhancer annotated in green. Bottom – Immunofluorescence image of whole-mount 23GFP E10.5 AGM, showing a single 2.5 μm slice. 23GFP marked HE and IACs. **(B)** Schematic of EHT cell populations sorted from E8.5, E9.5 and E10.5 mouse embryos, that were profiled by RNA-seq and ATAC-seq. Bars indicate population sort markers. Populations were first gated on Ter119$^-$. RNA-seq $n = 3$ (pre-HE = 4) and ATAC-seq $n = 4$ (pro-HSC = 3). *RNA-seq and ATAC-seq data were generated by previous members of the de Bruijn lab (G. Swiers and L.Greder, respectively) and were used in analyses and GRN construction performed by me.*

E10.5 mouse embryos (Fig. 3.1B, Appendix A.11A). These populations include E8.5 endothelial cells (EC, Ter119$^-$VECad$^+$CD41$^-$CD45$^-$23GFP$^-$) and pre-HE (Ter119$^-$VECad$^+$CD41$^-$CD45$^-$23GFP$^+$), E9.5 HE (Ter119$^-$VECad$^+$-CD41$^-$CD45$^-$23GFP$^+$) and pro-HSC (Ter119$^-$VECad$^+$CD41$^+$CD45$^-$), and E10.5 pre-HSC I (Ter119$^-$VECad$^+$CD41$^+$CD45$^-$) and pre-HSC II (Ter119$^-$VECad$^+$CD41$^+$CD45$^+$). Sort panels for these populations did not include CD43, so the population indicated as 'pro-HSC' contains both pro-HSC (CD43$^-$, Rybtsov *et al.*, 2014) and blood progenitors (CD43$^+$). E9.5 Ter119$^-$VECad$^+$CD41$^+$CD45$^-$ PAS cells contain 80.3 ±4.8% CD43$^+$ haematopoietic progenitors (Appendix A.11B).

Following pre-processing and mapping, reads were found to have high levels of PCR duplication (RNA-seq: 28% - 72.7% PCR duplication, with mean of 49.5%. ATAC-seq: 58.3% - 89% PCR duplication, with mean of 78.5%), and after PCR duplicate removal mapping was successful (Appendix A.3). Low-expressing genes were removed from the RNA-seq data (see methods section 2.14.1, p.50), before pre-processing with EdgeR (Robinson *et al.*, 2010). Peak-calls were generated from individual ATAC-seq samples, and peaks overlapping in at least two replicates were retained, in order to discard low-quality peak calls. ATAC-seq signal was quantified over these robust peaks to create a peak read count table, and as with the RNA-seq data low count peaks were discarded (see methods section 2.14.2, p.52).

3.2.1 Different EHT expression signatures drive differentiation, morphological, and signalling processes

To explore gene regulation in EHT, I first analysed general gene expression and chromatin accessibility profiles. As pairwise analysis between populations does not incorporate changes across the whole EHT process, I instead performed an ANOVA-like test in EdgeR (Robinson *et al.*, 2010) to extract differentially expressed genes (DEGs) from the RNA-seq data, and differentially accessible elements (DAEs)

from the ATAC-seq data. This method tests for changes between any populations, and therefore identifies gene expression and ATAC accessibility generally associated with EHT. ANOVA tests identified 5892 DEGs in the RNA-seq data, and 17969 DAEs in the ATAC-seq data.

PCA of the DEGs and DAEs clustered replicates together (Fig. 3.2A). The PC1 axis in both RNA-seq and ATAC-seq datasets represents a sequential progression of EHT. PC2 instead describes a transient activation or suppression of transcription,

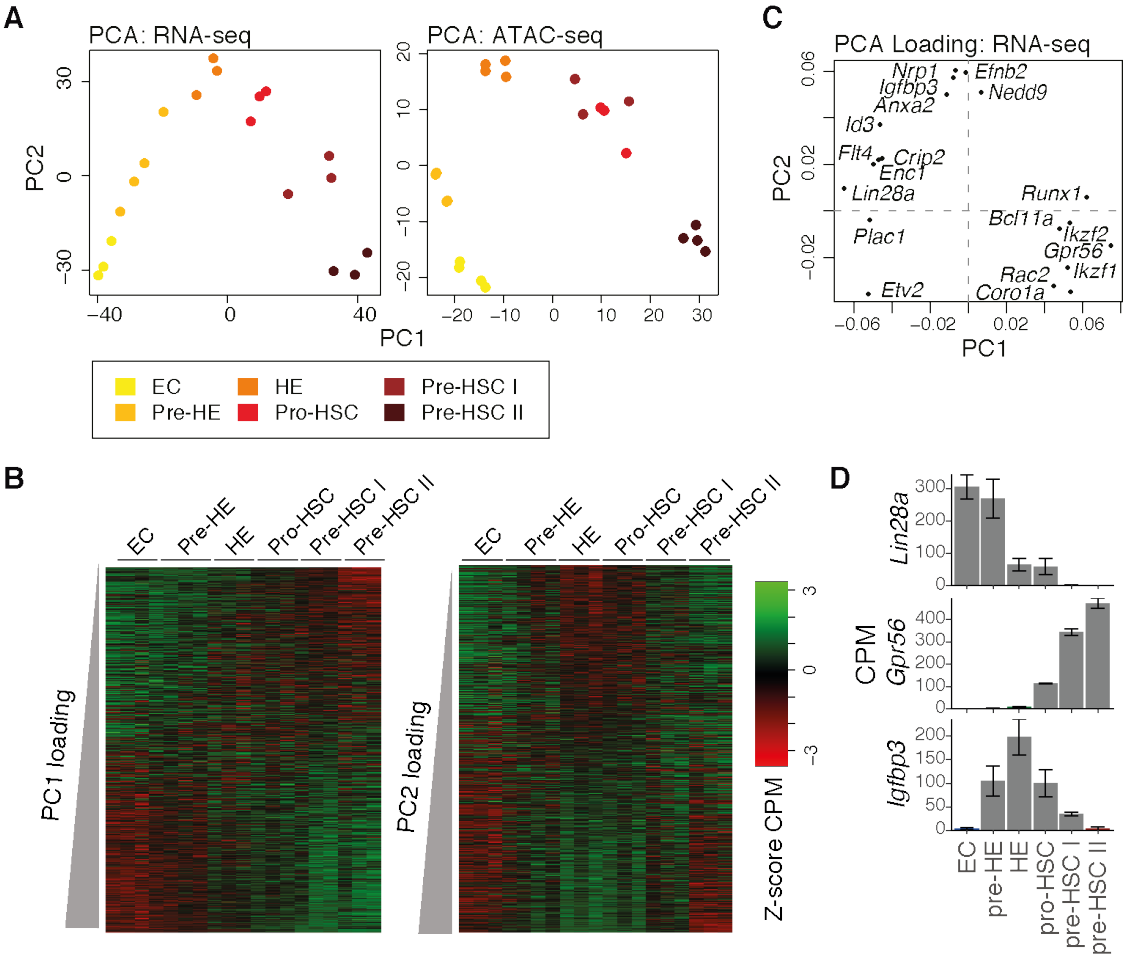


Figure 3.2: EHT principal components define switch and transient gene regulation. **(A)** PCA plots of RNA-seq (left) and ATAC-seq (right) samples that describe the EHT transition. ANOVA DEG and DAEs used as input. **(B)** Heatmap of RNA-seq gene expression (Log_2 CPM) ordered by PCA loading contribution to PC1 (left) and PC2 (right). **(C)** Top PC1+, PC1-, and PC2+ loadings relating to the RNA-seq PCA shown in A-B. **(D)** Average gene expression (CPM) plots for strong PC1 and PC2 loading genes. Error bars represent standard error of the mean; $n = 3$; $n = 4$ for pre-HE.

where expression peaks or troughs in HE/pro-HSC and returns to baseline levels in pre-HSCs (Fig. 3.2B). For example, *Lin28a* is PC1- associated with high EC expression and is known to negatively regulate haematopoiesis (Chaudhuri *et al.*, 2012), while *Gpr56* is PC1+ associated with high pre-HSC expression and is required for HSC generation (Solaimani Kartalaei *et al.*, 2015) (Fig. 3.2C-D). *Igfbp3* is PC2+ associated and transiently regulated with peak expression in HE (Fig. 3.2C-D), and is implicated in a range of pathways (Baxter, 2013) but has not previously been implicated in EHT. PC1 loadings can be considered to reflect genes that switch on or off, while PC2 loadings represent transient regulation.

To properly categorise these DEGs and DAEs into different patterns, the RNA-seq and ATAC-seq counts were clustered into five regulatory modules (RNA1-5 and ATAC1-5, Fig. 3.3A-B). The modules represent transient decrease (RNA1/ATAC1) or increase (RNA3/ATAC3) in RNA-seq and ATAC-seq counts, an endothelial bias (RNA2/ATAC2), early haematopoietic commitment (upregulation in HE/pro-HSC, RNA4/ATAC4), and late haematopoietic commitment (upregulation in pre-HSC I/II, RNA5/ATAC5). It is striking that similar patterns are observed between RNA-seq and ATAC-seq datasets. Clusters have been labelled to represent similar patterns between the datasets, such as RNA2 and ATAC2, both of which are downregulated over EHT.

As enhancer and promoter activity should drive gene expression changes it is expected that ATAC-seq peaks would be associated with genes within the same module, such as ATAC2 accessible elements linked to RNA2 genes. To test this hypothesis, I annotated each DAE to the closest DEG to associate enhancer and promoter elements to the corresponding gene. I then performed Fisher's exact tests to determine the significance of overlap between genes in RNA1-5 and genes linked to accessible elements in corresponding ATAC1-5 modules (Fig. 3.3B). Each corresponding RNA-seq and ATAC-seq module, other than RNA1 and ATAC1, was significantly associated with each other, supporting the hypothesis that differential enhancer and promoter accessibility is driving gene expression changes.

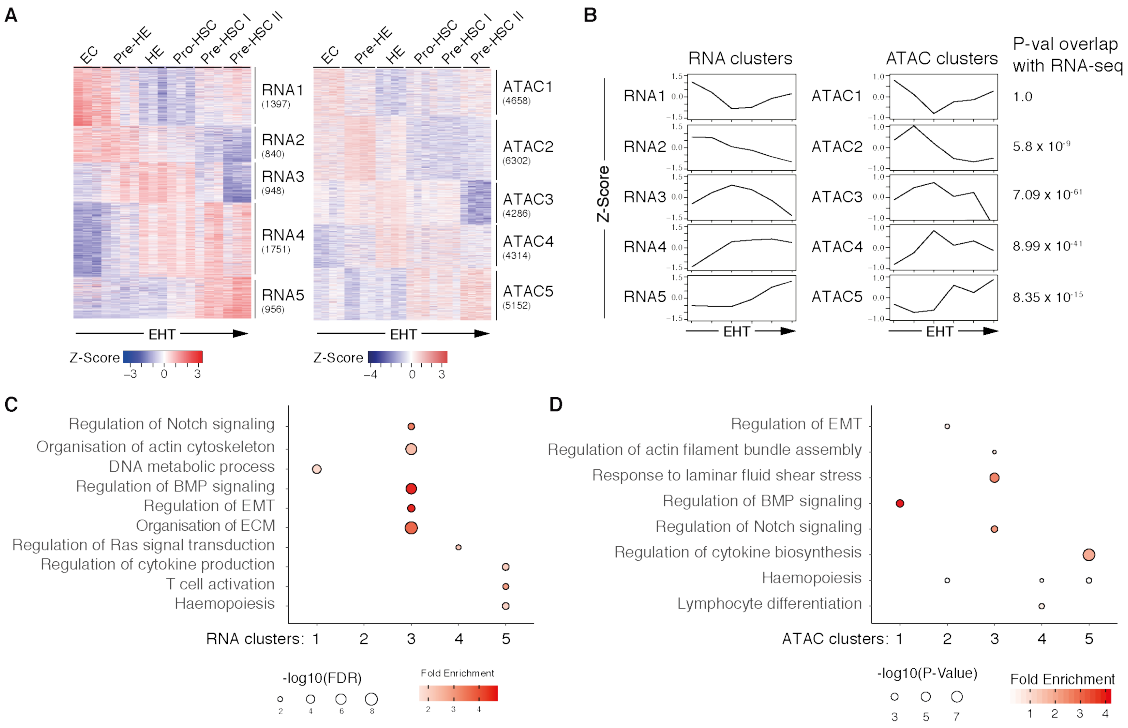


Figure 3.3: Transient regulation activates signalling and morphological changes while switches drive a haematopoietic phenotype. (A) Expression and accessibility heatmaps (Z-score Log_2 CPM) for RNA-seq DEGs (left) and ATAC-seq DAEs (right). Expression grouped by hierarchical clustering, and accessibility by *k*-means clustering. **(B)** Average z-score Log_2 CPM of gene expression (left), or chromatin accessibility (right) by clusters as in A. Fishers exact test results comparing the overlap of corresponding RNA- and ATAC-seq clusters are shown on the side. **(C-D)** GO biological process enrichment for modules RNA1-5 (C) and ATAC1-5 (D). Only significant data points shown ($P \leq 0.05$).

These modules describe a range of dynamic profiles across EHT, though it is unclear whether clustering the data has functional relevance. To determine whether these modules drive specific programs, I performed gene ontology (GO) enrichment analyses (Fig. 3.3C-D). Transiently activated genes (RNA3/ATAC3) are associated with Notch and BMP signalling, and processes that underpin morphological changes and migration, including epithelial-to-mesenchyme transition (EMT) and actin cytoskeleton organisation (Fig. 3.3C-D). Late haematopoietic commitment (RNA5/ATAC5) is instead enriched for terms related to cytokine production, haematopoiesis, and T cell activation, corresponding with a differentiation program. This breakdown of regulatory patterns highlights the complex regulation over EHT and begins to categorise distinct regulatory programs. In particular, transient

activation is associated with signalling and morphological changes that may drive the budding process during EHT.

3.2.2 Transition-specific regulation

EHT is controlled through specific sequential transitions between cell types. Therefore, an initial characterisation of RNA-seq and ATAC-seq changes at these transitions is important to understand. In particular, understanding the difference between E8.5 ECs and E8.5 pre-HE can identify key genes that distinguish this pre-HE population. To identify specific regulatory changes characterising cell type transitions I performed pairwise differential analyses between sequential EHT populations. As with the ANOVA analysis, DAEs were associated with the closest DEG promoter. As it is expected for changes in chromatin accessibility to reflect gene expression, I examined the relationship between RNA-seq and ATAC-seq logFC between transitions. Changes in chromatin accessibility and gene expression are generally concordant, such as with the EC to pre-HE transition where 79% of opening chromatin elements associate with upregulated genes (Fig. 3.4A). This concordance is not present in all transitions, such as the pre-HE to HE transition where 33% of closing elements are associated with increasing expression (Fig. 3.4A). This could be due to incorrect assignment of elements to gene promoters, or reflect silencers or non-regulatory elements.

To find regulatory elements associated with cell type transitions, I focused on strongly differential genes and accessible elements (Fig. 3.4B). I selected example genes and elements displaying among the highest or lowest logFC values between transitions, that were relevant to EHT based on known literature. The EC to pre-HE transition, marking the earliest initiation of EHT signalling, shows strong upregulation of *Cd109* expression and opening of a +4kb element, and a decrease in *Dlk1* and a -7kb element (Fig. 3.4B). CD109 is known to be expressed in CD34+ bone marrow cells, marking haematopoietic populations (Murray *et al.*, 1999) (though decreases in expression with HSC maturation in the embryo, see

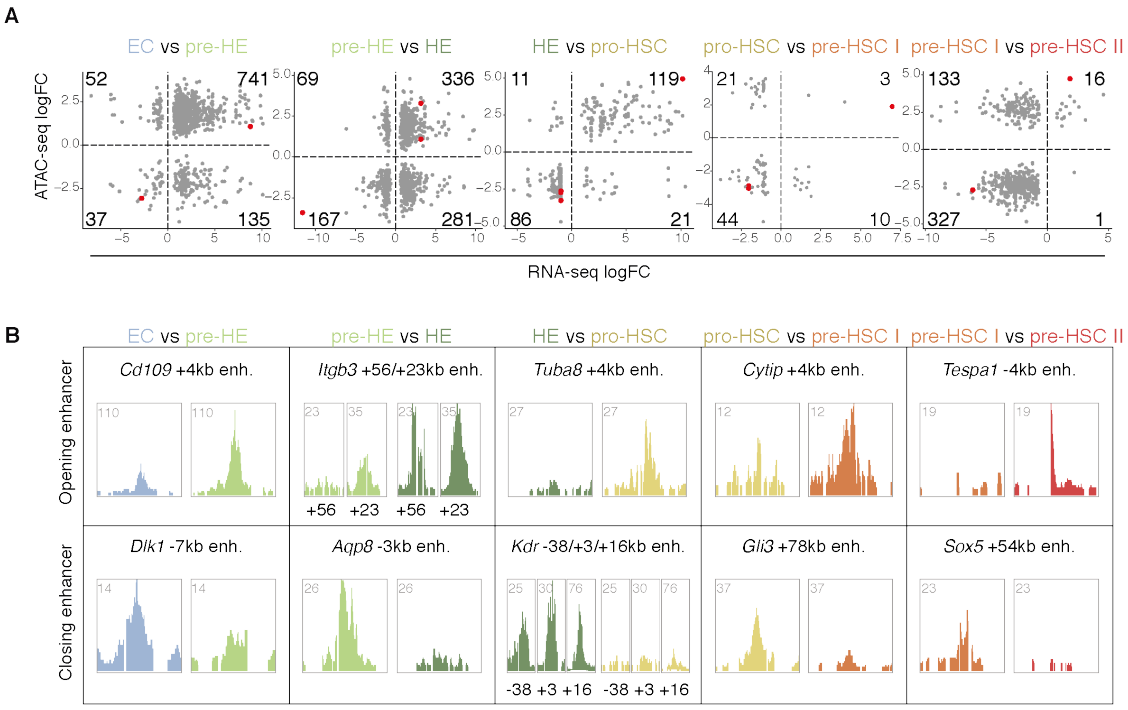


Figure 3.4: Analysis of enhancer-promoter pairs suggests stage-specific epigenetic regulation. (A) Relationship between RNA- and ATAC-seq logFC at significant pairwise comparisons. Each data point represents an accessible element paired with a gene (as defined by closest DEG promoter to the element). Only non-promoter elements are shown (distance > 1kb from TSS). High logFC genes that have relevance to EHT based on existing literature were highlighted red, and examined in B. **(B)** ATAC-seq tracks for DAEs, as highlighted in A. Sample replicates were combined into a single track. Values indicate track height. Opening elements: logFC > 0 and FDR < 0.05, closing elements: logFC < 0 and FDR < 0.05.

Fig. 3.5B). Dlk1 is a known negative regulator of EHT normally expressed in the surrounding AGM niche (Mirshekar-Syahkal *et al.*, 2013). Genes upregulated in the pre-HE to HE transition include *Itgb3* (CD61) which is a known marker of HE (Huang *et al.*, 2016). HE to pro-HE is characterised by deactivation of *Kdr* (Flk1), a receptor marking vascular endothelium (Millauer *et al.*, 1993), and increased *Tuba8*, a tubulin (Stanchi *et al.*, 2000), which reflects a loss of endothelial identity and cytoskeletal changes, respectively. Few genes are upregulated in pro-HSC to pre-HSC I, and pre-HSC I to II transitions, with the majority of genes being downregulated. Among downregulated genes in these transitions are several key TFs, including *Gli3*, which is known to modulate hedgehog signalling (Chaudhry *et al.*, 2017), and *Sox5* which modulates BMP signalling (Nordin *et al.*, 2014).

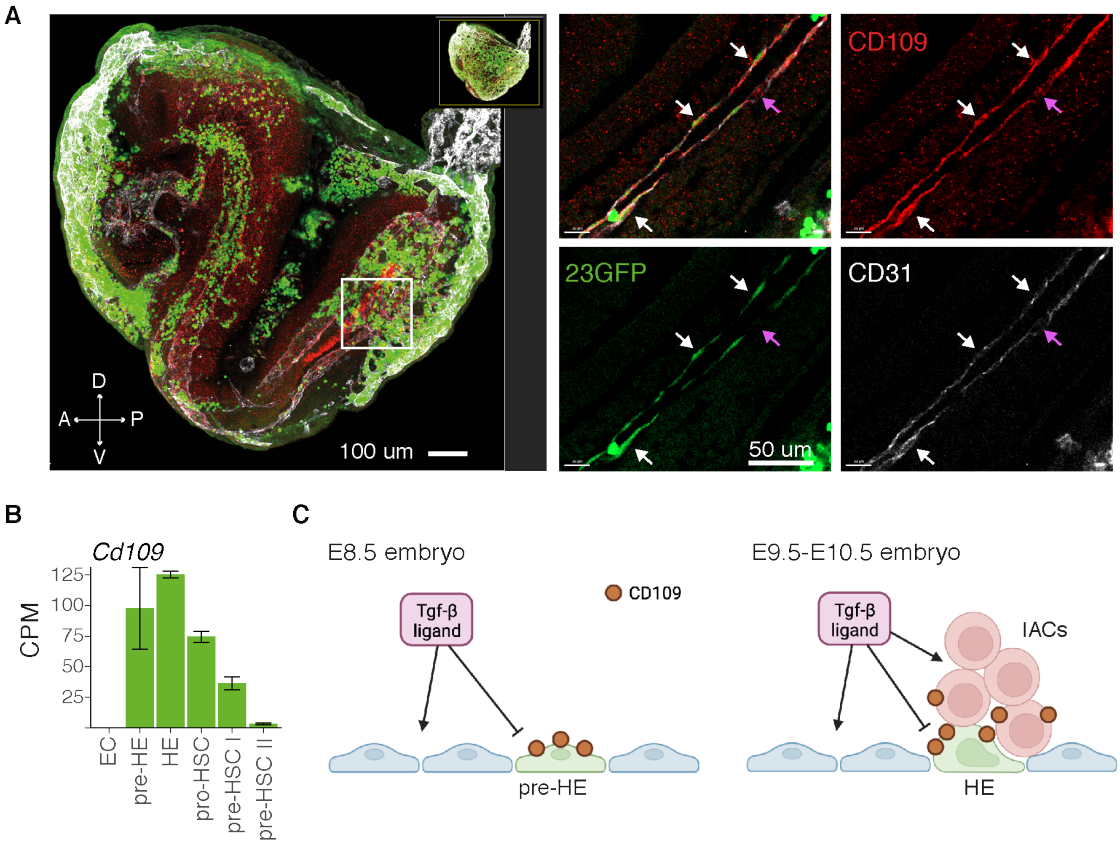


Figure 3.5: *CD109* expression and protein localisation in E8.5 embryos. (A) Confocal whole-mount immunofluorescence images of an E8.5 (5 somite pairs) 23GFP embryo. Left panel shows maximum intensity projection of paired dorsal aorta (225 μm thick Z-stack), with maximum project of whole embryo shown at the top right (340 μm thick Z-stack). Square box highlights region of interest. Right panels show a single 5 μm thick z-stack slice of the region of interest. White arrows indicate examples of CD109 and 23GFP overlap in endothelium, and purple arrows indicate an example of 23GFP-endothelium with low CD109. **(B)** Gene expression (CPM) for *Cd109* in bulk RNA-seq data. Error bars represent standard error of the mean; $n = 3$; $n = 4$ for pre-HE. **(C)** Schematic illustrating hypothesised activity of TGF-β, on the basis of CD109 repressing TGF-β receptor activation. *Panel C created with BioRender.com.*

The high expression of *CD109* in pre-HE is of particular interest, as CD109 is known to attenuate TGF-β signalling (Mii *et al.*, 2019). *Cd109* was expressed in E8.5 pre-HE, but not ECs, which was further validated by immunofluorescence imaging showing CD109 protein overlaps with 23GFP+ pre-HE in E8.5 embryos (Fig. 3.5A-B). Additionally, *Cd109* expression decreases as EHT progresses from pro-HSC to pre-HSC II (Fig. 3.5B). Together, this data highlights that CD109 activity is enriched in pre-HE and HE over other EHT populations. As CD109 may repress TGF-β activity, this suggests that the TGF-β pathway is inactive

specifically in pre-HE and HE, but maintains activity in E8.5 EC, and more mature pre-HSC populations (Fig. 3.5C). Maintained TGF-β activity in ECs is important, as the TGF-β pathway is a regulator of EC proliferation (Goumans *et al.*, 2002; Lebrin *et al.*, 2005). Further study is required to determine whether this CD109 enrichment is sufficient to impact TGF-β signalling, and whether CD109 has an effect on EHT initiation or progression.

3.3 Generating, and characterising a GRN model of EHT

To better understand the regulatory logic driving EHT, I constructed a GRN model using a multi-step approach to integrate RNA-seq and ATAC-seq profiles (Fig. 3.6A, Appendix A.9, methods section 2.15.2, p.56). Accessible regions were assigned to the closest active promoter, and those with strong correlation (> 0.4 or < -0.4) between ATAC-seq counts of the accessible element, and RNA-seq counts of the assigned promoter, were classified as enhancer-promoter associations (E-Ps). As both positive and negative correlation is incorporated, this definition of E-P also includes negative regulating elements (silencers). This definition of silencers has been used in Cornejo-Páramo *et al.*, 2022. While nearest promoter is not a robust definition, as enhancers can have low regulatory function or skip promoters to act on distal genes (Chepelev *et al.*, 2012), the correlation filter enriches for valid regulatory interactions (as used in Hariprakash *et al.*, 2019; Sheffield *et al.*, 2013). To focus on dynamic regulation, only E-Ps connecting differential enhancers with differential expression were retained (2810 DEGs and 6743 associated DAEs, Fig. 3.6B). Upstream regulators at enhancers and promoters were inferred by presence of TF binding motifs. To improve reliability, only high confidence motif matches were used ($P < 1 \times 10^{-6}$). Only robust interactions were retained, with strong RNA-seq correlation (> 0.4 or < -0.4) between the TF and predicted TF target. This established an integrated network of EHT regulatory interactions.

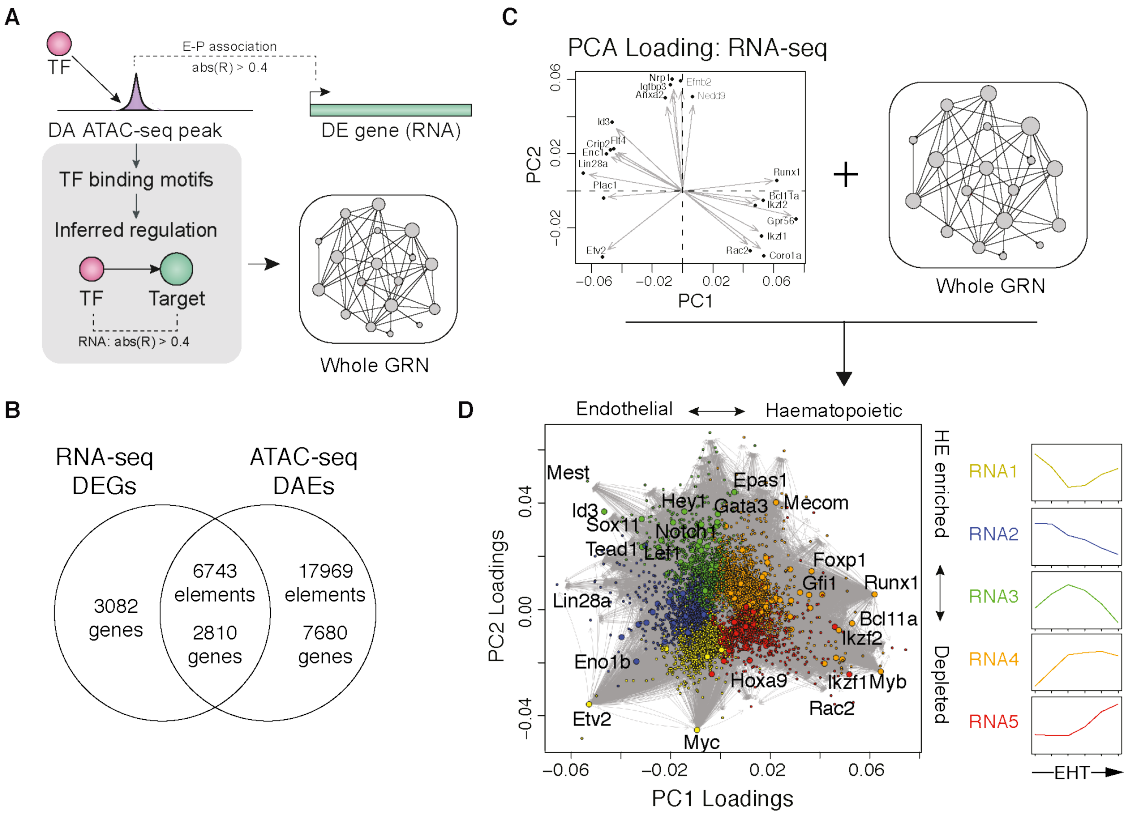


Figure 3.6: Constructing a GRN model of EHT. (A) Schematic outlining approach to generate the EHT GRN (methods section 2.15.2, p.56). **(B)** Overlap of RNA-seq DEGs and genes in E-Ps at DAEs. **(C)** Schematic illustrating approach to integrate RNA-seq PCA loadings (Fig. 3.2C) with the GRN model, to arrange the network nodes based on PCA loading coordinates (methods section 2.15.2, p.56). **(D)** Visualization of the EHT network embedded in RNA PCA loading coordinates (Fig. 3.2C), as illustrated in C. Nodes are coloured according to RNA-seq module (Fig. 3.3A-B). Module expression illustrated (right). All TF nodes are enlarged, with select TFs labelled.

To interpret this model in the context of EHT, the network was embedded into RNA-seq PCA loading coordinates (Fig. 3.2C) and annotated by expression modules (Fig. 3.3A). This stratified nodes based on expression profiles, describing EHT progression (PC1) and transiently expressed genes (PC2) (Fig. 3.6C-D). For example, this clearly marks major nodes of the mature haematopoietic program (RNA4 and RNA5), including *Gpr56*, *Bcl11a*, *Rac2*, and *Myb*. The transient program (RNA3) was instead characterised by signalling associated genes, including *Tead1*, *Sox11*, *Gata3*, and multiple Notch pathway genes. Together, this establishes an integrated model of EHT regulatory interactions, that organises the primary drivers of the five identified expression modules.

3.3.1 Central regulators and targets of the EHT GRN

To identify key regulators of the five expression modules, I calculated centrality measures to describe GRN nodes based on their connections within the network. Degree centrality is a measure of how many interactions are connected to a node (methods section 2.15.5, p.59), and in other systems has shown to correlate with gene essentiality (Hahn *et al.*, 2005; Jeong *et al.*, 2001; Koschützki *et al.*, 2008). Within the early haematopoietic commitment module (RNA4) the TFs with the highest degree centrality include Sp4, Smad3, and Klf4, in addition to the expected TFs Runx1, Gfi1, and PU.1 (*Spi1*) (Fig. 3.7A). Late haematopoietic commitment (RNA5) central nodes include Elf1, Nr1d2 and Ikzf1. TFs central to the transient module (RNA3) include Ets2, Nfib, Lef1 and Gata3/6. In addition to highly central TFs, genes with many accessible enhancers may indicate loci highly regulated during EHT (Fig. 3.7B). *Kdr* (Flk1) and *Aplnr*, a marker of endothelium previously suggested to give rise to HE (Crosse *et al.*, 2020), have many endothelium biased enhancers (ATAC2). Among transiently accessible enhancers *Igfbp3* is associated with many accessible enhancers (Fig. 3.7B) and drives HE identity (Fig. 3.2C-D). The central TFs *Nfib* and *Fli1* are also regulated by several transient enhancers. Among haematopoietic expression modules (RNA4 and RNA5) *Ikzf1*, which is a highly central RNA5 TF (Fig. 3.7A), has many accessible enhancers (ATAC5). The frequency of enhancers at these predicted key drivers of EHT, suggests that important factors are robustly regulated through multiple enhancers.

3.4 *Runx1* regulators and target genes

As Runx1 is a critical regulator of EHT, it is important to understand its role in the GRN. To begin to investigate how Runx1 drives gene expression changes, it is important to understand the regulatory context of *Runx1*. Specifically, I explored novel upstream regulators of *Runx1*, and downstream targets coinciding with Runx1 activity.

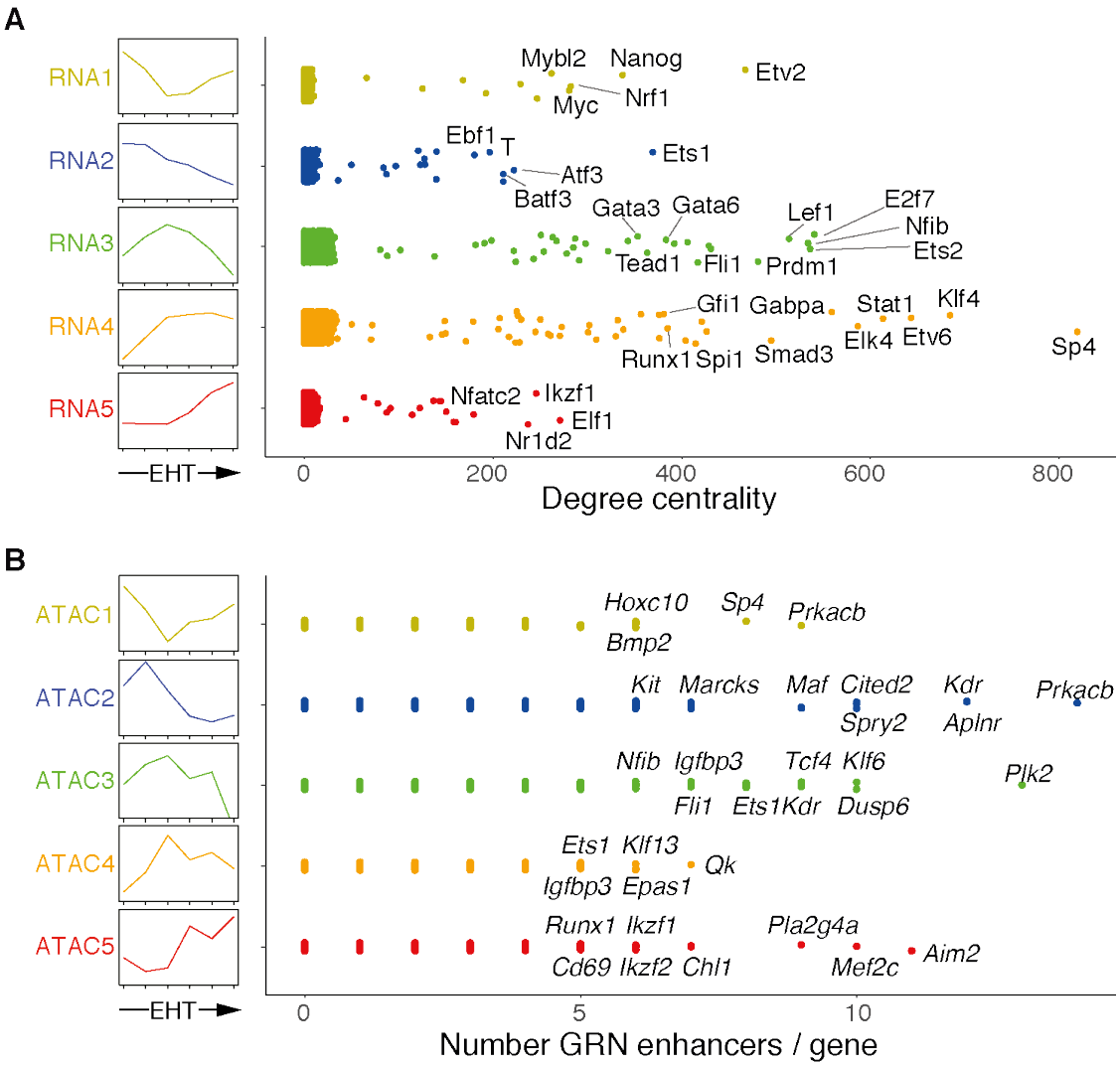


Figure 3.7: Centrality analysis predicts TF drivers of EHT modules. (A) Distribution of degree centrality of GRN nodes, stratified by RNA-seq expression clusters (RNA1-5). Select genes have been labelled to highlight top central regulators and key haematopoietic TFs. **(B)** Number of enhancers linked to each gene, stratified by ATAC-seq accessibility clusters (ATAC1-5). Select genes shown to highlight example loci with many enhancers.

3.4.1 Cis-regulatory landscape of the *Runx1* locus

To predict potential regulators of *Runx1*, it is necessary to first establish the cis-regulatory landscape of the *Runx1* locus that may drive regulation. To accomplish this, I assembled a detailed map of in vivo chromatin accessibility over the *Runx1* locus based on ATAC-seq data (Fig. 3.8). The EHT GRN detects several enhancers correlating with *Runx1* expression, with the strongest being +23 ($R > 0.9$). These

include previously published enhancers with -371, -354, -59, -48, +3, +23, +24, +110, and +204 regulatory elements (as summarised in Owens *et al.*, 2022), and the +38 element which has not previously been published. The +24 element is a second peak of accessibility adjacent to the +23 enhancer, that boosts +23 activity (de Bruijn lab, unpublished).

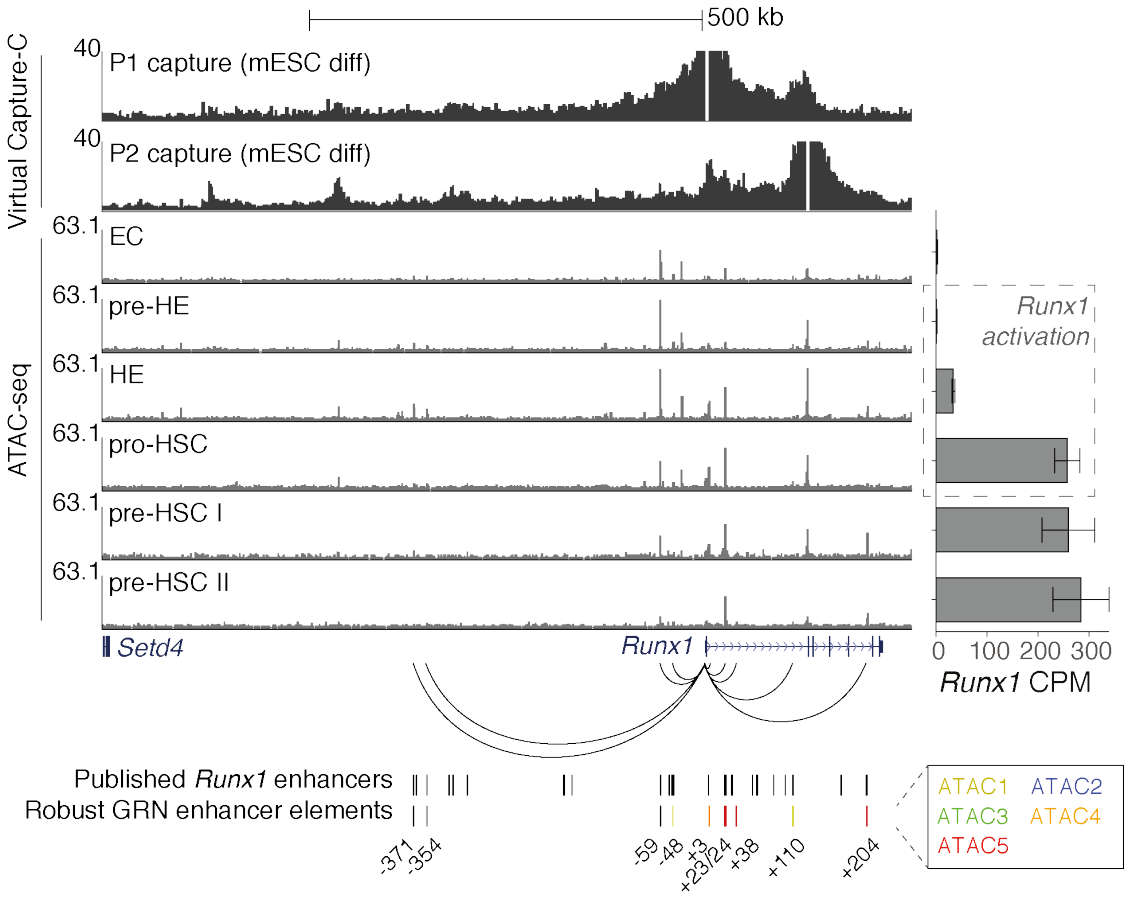


Figure 3.8: Chromatin accessibility profiles and predicted enhancers across the *Runx1* locus. ATAC-seq and "virtual" capture-C tracks over the *Runx1* locus and upstream gene desert. Virtual capture-C tracks were taken from tiled-C data performed on mESC differentiated derived $CD41^{+}$ HPCs, using the P1 and P2 promoters as anchor points (Owens *et al.*, 2022), to show the frequency of physical interactions between promoters and surrounding DNA. For ATAC-seq tracks, replicates were combined into a single track for visualisation. GRN predicted enhancers and published enhancers (Bee *et al.*, 2010; Cauchy *et al.*, 2015; Cheng *et al.*, 2018; Fitch *et al.*, 2020; Harland *et al.*, 2021; Marsman *et al.*, 2017; Ng *et al.*, 2010; Nottingham *et al.*, 2007; Ortt *et al.*, 2008; Owens *et al.*, 2022; Schütte *et al.*, 2016) were annotated, coloured by accessibility profiles (ATAC1-5). *Runx1* expression shown on right. Error bars represent standard error of the mean; $n = 3$; $n = 4$ for pre-HE. *Tiled capture-C data was generated and analysed by D. Owens (Owens* et al.*, 2022).*

Of these predicted regulatory elements, -59, -48, and +23/+24 are accessible in ECs (though +23 low signal prior to HE could be due to presence of the +23GFP transgene). Low level -371 and -354 accessibility is initiated in pre-HE and increased in HE, then closed from pro-HSCs onwards (Fig. 3.8). From pre-HE to HE, the +3, +23 and +24 enhancers increased accessibility. +23 and +24 remain open throughout EHT, and +3 remains accessible until closing in pre-HSC II. The -54 and -48 enhancers decreased in accessibility as EHT progressed from pro-HSCs onwards, suggesting they are silencers of the locus, or that they are only involved in initial *Runx1* induction. The +38 enhancer is weakly accessible in pre-HSC populations. The +204 element is weakly accessible in pre-HE and HE, closed in pro-HSC, but strongly accessible in pre-HSC populations. Additionally, the +110 enhancer has weak activity specifically in the pro-HSC population. This increased accessibility of +110, and absence of +204 accessibility, in the pro-HSC population is likely to reflect the presence of CD43+ HPCs, as discussed in section 3.2.1 (de Bruijn lab, unpublished).

To evidence that these GRN predicted elements regulate *Runx1*, I took tiled capture-C analysis previously performed on mESC haematopoietic differentiation populations ($Flk1^+$ mesoderm and $CD41^+$ HPCs) in the de Bruijn lab (D. Owens, Owens *et al.*, 2022). These predicted regulatory elements displayed physical interaction with either the P1 or P2 promoters, though -371 and -354 showed low interaction frequency (Fig. 3.8). Analysis from Owens *et al.*, 2022 showed -59, -48, +3, +23, and +110 enhancers to increase in P1 and P2 interactions when mESCs are differentiated from mesoderm to HPCs. Together, these analyses and observations construct a comprehensive map of *Runx1* regulatory enhancers and silencers that drive *Runx1* expression.

3.4.2 Predicted regulators of *Runx1* expression

To predict potential TFs driving *Runx1* expression I aimed to explore TF binding motifs across the cis-regulatory landscape. There are three cell population transitions

of particular relevance to *Runx1* regulation: the EC to pre-HE transition marked by +23GFP expression (Swiers *et al.*, 2013a), the pre-HE to HE transition marked by initiation of endogenous *Runx1* expression, and the HE to pro-HSC transition where *Runx1* expression is upregulated (Fig. 3.8). I interrogated potential drivers of *Runx1* transcription by exploring TF binding motifs present at the enhancers of E-P interactions (Fig. 3.9A-B, see methods section 2.15.1, p.56). Only high confidence TF motif PWMs from the HOCOMOCO motif database were used, to improve confidence of predicted regulatory interactions (core HOCOMOCO, Kulakovskiy *et al.*, 2018). Motifs were stratified by pairwise differential analyses between populations to predict transition-specific drivers.

To characterise the factors that may establish the pre-HE state in E8.5 embryos, I first probed TFs upregulated from EC to pre-HE, with motifs in +23/24 enhancers. These include TFs associated with different motif families, such as ETS (Elk4), FOX (Foxo1), zinc fingers (Klf6, Prdm16 and Sp4), Maf-related bZIP (Maf), NFI (Nfib), SMAD (Smad3), STAT (Stat1) and basic Helic-Loop-Helic (bHLH) E2A (Tcf4) (Fig. 3.9A). Among these differentially expressed TFs, *Maf* expression is most specific to pre-HE (Fig. 3.9B). While several of these TFs are reported targets of Runx1 (Wang *et al.*, 2011; Zaidi *et al.*, 2002), their expression precedes *Runx1* and coincides with 23GFP activity, suggesting an additional upstream regulatory role. Elements open in pre-HE, other than +23/24, include -371, -354, and +204 at low levels of accessibility, as well as -59 and -48 at high levels of accessibility (Fig. 3.8). There are a number of TFs upregulated from EC to pre-HE, with predicted motifs in these enhancers. These include motif families such as GATA (Gata3), STAT (Stat4), Rel homology domain (RHD) NF-κB (Nfkb1), three amino acid loop extension (TALE) homeobox MEIS (Meis1, Meis2) and PBX (Pbx1), Ikaros zinc finger (Ikzf1), and Creb-related bZIP (Crem). GATA motifs are found in -371, -354, and -59 elements, coinciding with *Gata3* upregulation. Gata3 has recently been reported as essential in aortic endothelium for EHT progression and promotes *Runx1* activity (Zaidan *et al.*, 2022), and therefore may be important for the pre-HE population.

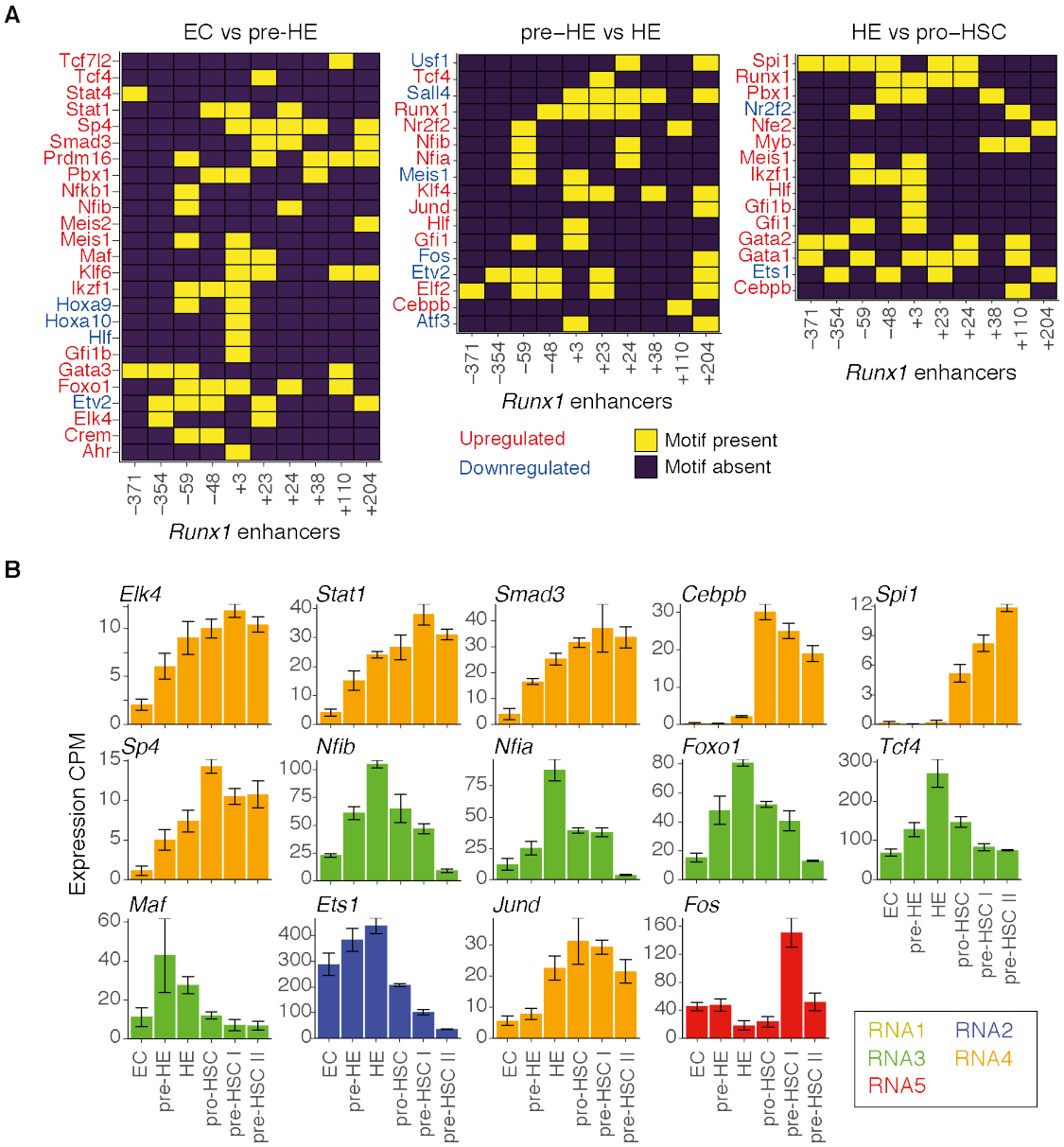


Figure 3.9: Predicted upstream regulation of *RUNX1*. **(A)** Heatmap of TF binding motif presence ($\geq$ 1 motifs) at *Runx1* enhancers. Motifs match TFs differentially expressed over the relevant pairwise comparisons, as indicated. Row labels refer to TFs associated with a detected motif, and coloured according to up- or downregulation over the transition. Columns refer to the individual *Runx1* enhancers. **(B)** Gene expression (CPM) for predicted *Runx1* enhancer regulators. Error bars represent standard error of the mean; $n = 3$; $n = 4$ for pre-HE.

Initiation of endogenous *Runx1* transcription (pre-HE to HE) and *Runx1* upregulation (HE to pro-HSC) are dynamic stages likely induced by key TFs. NFI motifs (Nfia and Nfib) underlie the -59 and +24 enhancers, and *Nfia* and *Nfib* genes are most expressed in HE (Fig. 3.9A-B). Nfib was confirmed to bind

to the +24 enhancer (but not the -59 element) using published ChIP-seq data in mammary cells (Appendix A.12, Shin *et al.*, 2016). CEBP motifs were unique to the +110 enhancer, and the *Cebpb* gene upregulated in both transitions. Cebpb binding to the +110 enhancer was validated using published ChIP-seq in pre-B cells (Appendix A.12, van Oevelen *et al.*, 2015). Similarly, we find AP-1 motifs (Fos and JunD) motifs underlying the +204 enhancer, and the *Fos* gene is upregulated in the pre-HE to HE transition. Further, the binding of Fos and JunD to the +204 enhancer was confirmed in macrophages and T-cells using published ChIP-seq data (Appendix A.12, Eichenfield *et al.*, 2016; Roychoudhuri *et al.*, 2016). These observations suggest Cebpb, Fos and Jund as unique regulators of the +110 and +204 enhancers, respectively. *Cebpb* upregulation occurs from HE to pro-HSC, while *Jund* is upregulated from pre-HE to HE, potentially placing Cebpb and Jund as transition-specific drivers. Additionally, Nfia and Nfib are potential novel regulators of the +24 enhancer. It is important to acknowledge that motif analyses based on PWMs are inherently predictive, and while broadly accurate genome wide, may not correctly identify motifs at individual loci. Nevertheless, these analyses provide an interesting preliminary characterisation of potential novel regulators of *Runx1* enhancers.

3.4.3 Predicted targets of Runx1

The GRN model predicts Runx1 to regulate many gene targets. An important consideration of these predictions is that RUNX motifs were detected based on a PWM that extends beyond the core consensus motif (Kulakovskiy *et al.*, 2018), and as such likely overestimates predicted Runx1 targets. Despite this caveat, this model offers an exploration of potential Runx1 TF targets. To find key Runx1 targets that may drive EHT, I explored TFs that are differentially expressed between cell type transitions. Direct Runx1 targets are likely to be among those that change in expression just after *Runx1* is upregulated, which is best represented by the pre-HE to HE, and HE to pro-HSC transitions. The GRN model in these transitions detects known Runx1 targets, such as *Spi1* (PU.1) (Huang *et al.*, 2008), *Gfi1* (Wilson *et al.*, 2010), *Nfe2* (Wang *et al.*, 2010) and *Ikzf1* (Iacovino *et al.*,

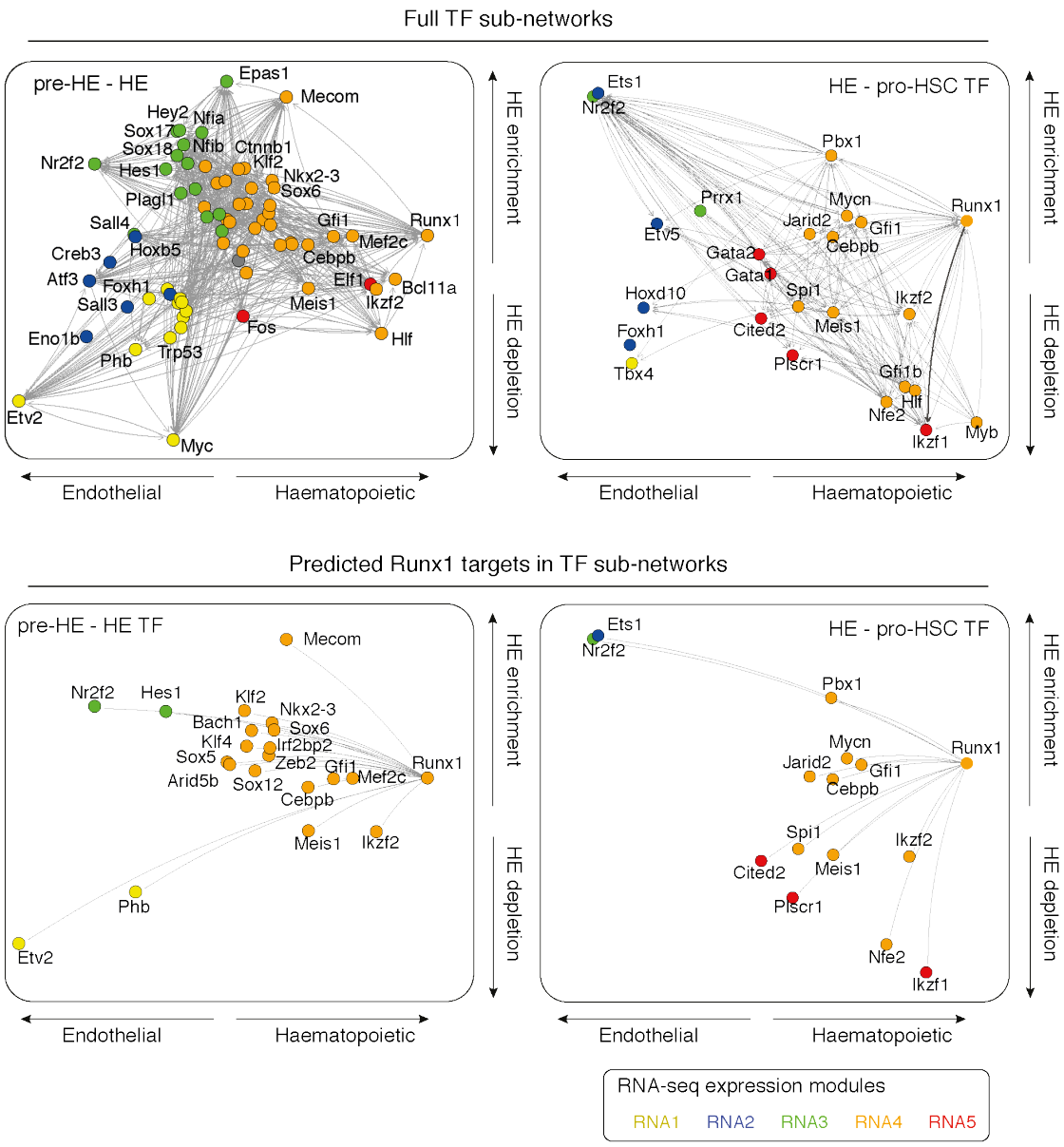


Figure 3.10: Transition specific TF sub-networks highlight predicted Runx1 targets. (A) Sub-networks of EHT GRN (Fig. 3.6C) illustrating up and downregulated TFs between pre-HE and HE (left) or HE and pro-HSC (right). Nodes are coloured based on expression modules (RNA1-5). Top sub-networks show all differential TFs, and bottom sub-networks show TFs predicted to be regulated by Runx1.

2011) (Fig. 3.10). *Zeb2* is an interesting target that was recently found to be a direct target of Runx1 in a murine kidney cell line, where expression of *Zeb2* in *Runx1* knockout models was sufficient to partially restore the transcriptional phenotype of Runx1 (Hass *et al.*, 2021). Zeb2 is a known EMT regulator that represses cell-cell junction genes and promotes haematopoietic differentiation (Li

et al., 2017; Vandewalle *et al.*, 2005), and as such may drive the morphological changes that underlie EHT. *Cebpb* is also predicted to be a Runx1 target, and has been previously shown to regulate several key haematopoietic TFs, such as *Gata2* (Goode *et al.*, 2016). As shown in Fig. 3.9, Cebpb may regulate the +110 *Runx1* enhancer, resulting in a possible positive feedback loop.

To find key Runx1 targets that may drive EHT, I examined predicted targets that are highly central, as analysed in Fig. 3.7A. *Etv2* is a highly central TF in the RNA1 module, with biased expression towards endothelium, that is predicted to be regulated by Runx1 (Fig. 3.10). Interestingly, *Etv2* is expressed in mesoderm and drives both endothelial and haematopoietic programs (Liu *et al.*, 2015; Menegatti *et al.*, 2019), and acts as a precursor of HE and EC fate. *Runx1* expression defines the haematopoietic fate of Etv2+ endothelial progenitors (Eliades *et al.*, 2016), and from this analysis may directly repress *Etv2*, in line with the known role of Runx1 to repress the endothelial program (Lancrin *et al.*, 2012). Ikzf1 is a prominent central TF in the RNA5 module (late haematopoietic expression) (Fig. 3.7A), and is a GRN predicted, and known target of *Runx1* (Fig. 3.10), as has been previously reported in mESC derived EBs (Iacovino *et al.*, 2011). Ikzf1 is expressed in multiple haematopoietic lineages, and is critical for B cell development (Georgopoulos, 2017; Marke *et al.*, 2018; Nichogiannopoulou *et al.*, 1999).

This analysis has predicted several known, and novel targets of Runx1. The role of several of these targets, such as Ikzf1, Zeb2, and Cebpb, is not fully understood. It remains a question as to whether Runx1 regulation is important for their normal function and contribution to EHT.

3.5 Runx1 cooperates with TFs to drive EHT processes

The analysis of the EHT GRN thus far has identified regulatory modules underlying key transitions (section 3.2.1, Fig. 3.3), the candidate drivers of each

module (section 3.3.1, Fig. 3.7), as well as the regulatory context of Runx1 (section 3.4, Fig. 3.8, Fig. 3.9, and Fig. 3.10). An important consideration in gene regulation is that TFs often do not act independently. While some TFs may act as pioneer factors in specific contexts (Zaret, 2020; Zaret *et al.*, 2016) and so can function independent of other factors, it is also common for TFs to act in complex with other proteins. The bZIP class of TFs is a classic example of dimerising proteins that modulate TF behaviour (Miller, 2009; Rodríguez-Martínez *et al.*, 2017). Runx1 is also known to cooperate with multiple other TFs (Goetz *et al.*, 2000; Hu *et al.*, 2011; Kim *et al.*, 1999b; Wang *et al.*, 2011; Zaidi *et al.*, 2002), and the expression of *Runx1* leads to genome-wide modulation of Fli1 DNA binding affinities (Lichtinger *et al.*, 2012). TF cooperation is challenging to explore on a genome-wide scope, but a study of how TFs may cooperate in EHT may offer greater insight into the regulation of EHT.

Here I aimed to apply the GRN model to predict potential cases of TF cooperativity. To predict TF co-interaction in the EHT GRN, I designed an approach to find TF pairs that are predicted to commonly regulate the same enhancer or promoter element. Note that in this thesis I am using "cooperativity" to refer to TFs that function synergistically, and "co-interaction" to refer to TFs predicted to regulate the same sites, as without additional experiments, synergy cannot be proven using the GRN model alone. I performed Fisher's exact tests for all possible TF pairs, in which a positive odds ratio describes TFs that co-interact more frequently than if interactions were random. Odds ratio scores were then grouped into five co-interaction clusters (CoInt 1-5, Fig. 3.11A). These clusters represent TFs that are predicted to co-interact at the same loci, and by GO analyses each cluster is enriched for different processes (Fig. 3.11B). Runx1 is in CoInt-1, which is enriched for multiple haematopoietic lineage differentiation processes, in line with driving a haematopoietic program (Fig. 3.11B). CoInt-2 is related to the stress response, and CoInt-3 involves cell cycle genes and responses to external factors. Interestingly,

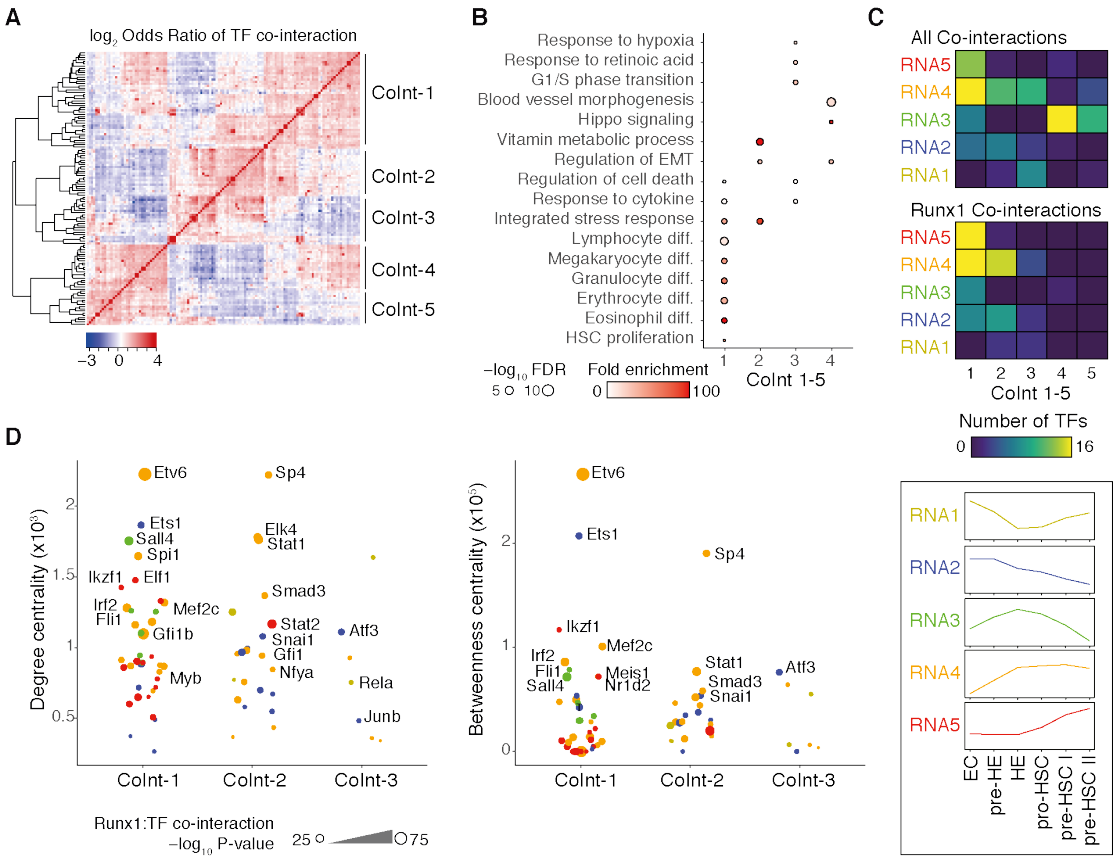


Figure 3.11: TFs cooperate in clusters to drive different EHT processes. (A) Hierarchical clustered heatmap of odds ratios resulting from Fishers Exact tests comparing the overlap of TF-targets (enhancers) between pairs of TFs. Benjamini and Hochberg correction of P values was performed, and TFs not significant for any interaction were excluded. Clusters describe sets of TFs that commonly act at the same loci, suggesting cooperative regulation. **(B)** GO biological process enrichment for clusters CoInt 1-4, as in A. Only significant data points shown ($P \leq 0.05$). Note that no terms were significantly enriched for CoInt-5 . **(C)** Frequency of TFs in each CoInt cluster, as in A, stratified by RNA expression modules (RNA1-5). **(D)** Stratification of degree (left) and betweenness centrality (right) of the EHT GRN, stratified by CoInt clusters. Only TFs found to significantly co-interact with Runx1 are shown (FDR < 0.05).

CoInt-4 is associated with blood vessel morphogenesis and EMT, suggesting a role in modulating aortic endothelium and driving morphological changes.

TFs in CoInt1-3 are enriched in the GRN module RNA4, so are activated in HE or pro-HSCs (Fig. 3.11C). CoInt-4/5 TFs are instead enriched in the transient expression module (RNA3), and likely represent a cooperative program that drives the effects of the transient module. This is emphasised by CoInt-4 driving EMT processes, in line with RNA3 association with EMT and actin cytoskeletal

organisation. Interestingly, when exploring *Runx1* expression correlation with predicted Runx1 targets across GRN modules, *Runx1* is found positively correlated with RNA4-5, but is anti-correlated with RNA-3 (Appendix A.13). This suggests that Runx1 activity represses transiently expressed genes. Runx1 co-interactors are primarily found in CoInt-1 and CoInt-2 (RNA4 enriched), but not CoInt-4 or CoInt-5 (RNA3 enriched) (Fig. 3.11C), suggesting that while Runx1 may regulate the transient program (RNA3), it does not cooperate with transient TFs. Runx1 co-interacts with multiple highly central TFs (Fig. 3.11D), including factors known to cooperate with Runx1 such as Fli1 (Lichtinger *et al.*, 2012), PU.1 (Hu *et al.*, 2011), and Ets1 (Goetz *et al.*, 2000; Kim *et al.*, 1999b). Several TFs that co-interact with Runx1 also display high betweenness centrality, such as Etv6, Ets1, Ikzf1, and Sp4 (Fig. 3.11D, Fig. 1.1, p.12). Several of these are predicted targets of Runx1, including Ets1 and Ikzf1. These high betweenness centrality nodes may be critical for GRN pathways, and may cooperate with Runx1.

This analysis successfully identifies TFs that commonly co-interact co-interact at the same loci, and specifically explores TFs that co-interact with Runx1. However, it does not provide evidence for synergistic cooperativity. Further, the GRN model is based on defined populations from bulk RNA-seq and ATAC-seq, and the relationship between these TFs may differ on a single cell level. To address this, I reanalysed published scRNA-seq EHT datasets in human and mouse (Zeng *et al.*, 2019; Zhu *et al.*, 2020). To compare gene expression patterns to *Runx1* transcription, I performed pseudotime analysis and correlated gene expression using expression values smoothed along the pseudotime axis (Fig. 3.12A-B, methods section 2.14.5, p.55). This revealed CoInt-1 TFs can be either correlated or anti-correlated with *Runx1* expression (Fig. 3.12C-E). *Ets1*, *Zeb1* and *Snai1* were consistently anti-correlated with *Runx1*, which may suggest that these factors cooperate with Runx1 in a very specific window before downregulation. Other TFs were positively correlated with *Runx1* expression, with *Etv6*, *Spi1*, *Myb*, *Elf1*, and *Ikzf1* showing the highest correlation. This positive correlation suggests these

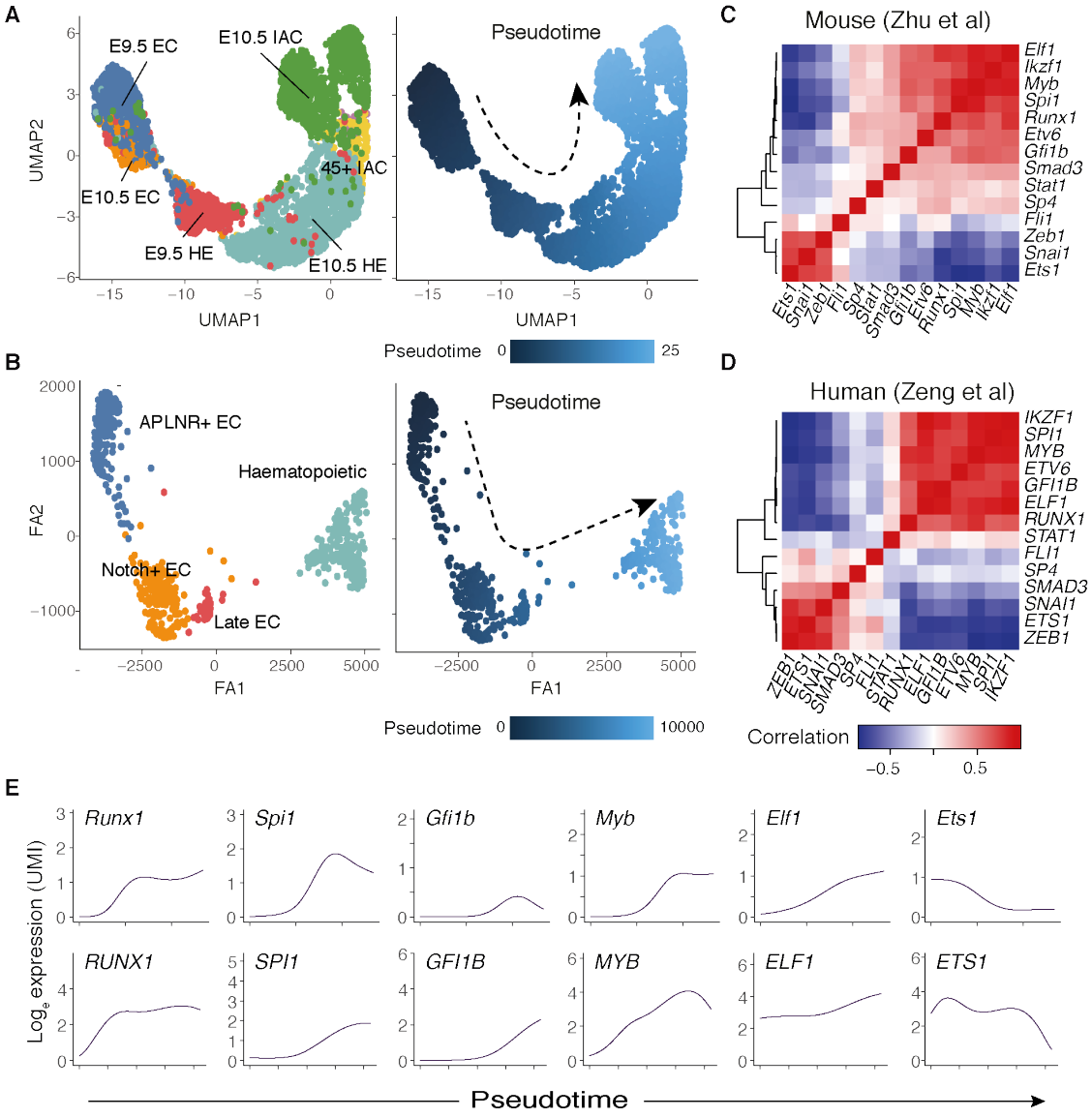


Figure 3.12: Gene expression correlation between CoInt-1 TFs in scRNA-seq data. (A, B) UMAP analysis of mouse EHT scRNA-seq from Zhu *et al.*, 2020 (A), and Force directed graph (ForceAtlas2, FA2) of human EHT scRNA-seq from Zeng *et al.*, 2019 (B). Sort populations labelled on left plots. Right plots show Slingshot pseudotime scores. Arrows represent direction of cell differentiation. **(C, D)** Correlation heatmaps of scRNA-seq expression from Zhu *et al.*, 2020 (C) and Zeng *et al.*, 2019 (D). See methods section 2.14.5, p.55 for single cell correlation calculation. **(E)** scRNA-seq expression of key Runx1 co-interacting TFs for mouse (top) and human (bottom) over pseudotime. Lines indicate smooth fitted values calculated by TradeSeq analysis. *Raw scRNA-seq data from Zhu* et al.*, 2020 and Zeng* et al.*, 2019, reanalysed by me.*

factors may synergistically cooperate, though further study is required. Overall, this analysis predicts significant TF co-interaction throughout EHT, and this cooperative logic may drive specific EHT processes.

3.6 Potential Runx1 driven FFLs in EHT

This framework to interrogate predicted TF cooperation allowed us to begin exploring how combinatorial regulatory logic may drive different regulatory processes. A common circuit in which TF cooperation occurs is through a FFL, where one TF regulates another TF, and both TFs cooperate to drive regulation of a downstream gene (section 1.2.6, p.13). FFLs are also highly enriched in biological networks (Mangan *et al.*, 2003a; Milo *et al.*, 2002), and are thought to mediate particular gene expression dynamics such as enacting sign-senstive delay in response to signals (Kalir *et al.*, 2005; Mangan *et al.*, 2003b), and acting as pulse generators (Basu *et al.*, 2004; Mangan *et al.*, 2003a).

Several TFs within CoInt-1 and CoInt-2 were found to frequently co-interact with Runx1, such as the high betweenness centrality TFs Etv6, Ets1, Sp4, and Ikzf1 (Fig. 3.11C-D). Additionally, a number of these co-interacting TFs were also found to be targets of Runx1, and in particular were differentially expressed subsequent to the onset of *Runx1* expression. These include Ets1, Ikzf1, Meis1, Nfe2, PU.1, Mef2c, Gfi1, Pbx1, and Cebpb (Fig. 3.13A, Fig. 3.10, Fig. 3.11), suggesting that Runx1 has the capacity to form cooperative FFLs with these TFs. I next probed what processes these co-interacting TFs may regulate. While the majority of CoInt-1 and CoInt-2 TFs are expressed in RNA4 and RNA5 modules (Fig. 3.11C), they may not necessarily interact with ATAC4 and ATAC5 accessible elements. To determine what ATAC-seq modules CoInt-1 and CoInt-2 TFs are most likely to regulate, I performed TF binding motif enrichment analysis across ATAC1-5 elements (Fig. 3.13A). For example, RUNX motifs are highly enriched in ATAC5 over other modules, suggesting Runx1 binds to and regulates pre-HSC-active genes. This analysis does not preclude Runx1 from regulating targets in other modules, and in-fact many of the TFs predicted to co-interact with Runx1 are instead associated with endothelial and transient accessible sites (ATAC2 and ATAC3). Runx1 may therefore cooperate with TFs across different ATAC-seq accessibility modules, in possible FFL configurations.

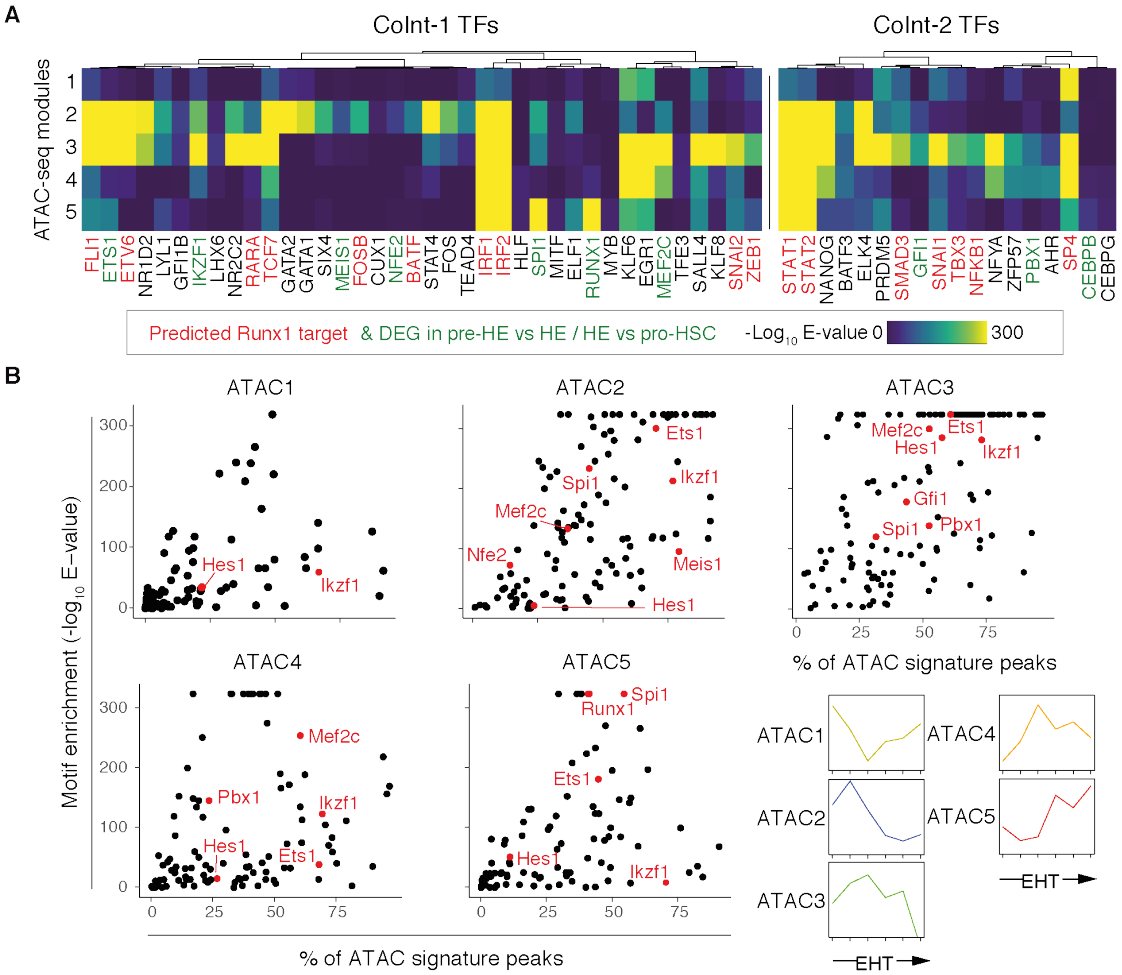


Figure 3.13: Motif enrichment of Runx1 co-interacting TFs across ATAC-seq accessibility modules. (A) Heatmap of TF DNA binding motif enrichment scores ($-\log_{10}$ E-value) (MEME-AME, methods section 2.15.1, p.56) within ATAC-seq elements found in modules ATAC1-5. Rows indicate peak sets from ATAC1-5 modules, columns indicate motif enrichment. Column labels refer to TFs (as in CoInt-1 and CoInt-2 co-interactions clusters) for which a motif PWM is associated, and used for enrichment analysis. TFs in CoInt-1 and CoInt-2 that co-interact with Runx1 shown. TF labels are coloured red if they are GRN predicted targets of Runx1, or green if they are predicted targets of Runx1 and are differentially expressed in either the pre-HE to HE, or HE to pro-HSC transitions (see Fig. 3.10). **(B)** Relationship between motif enrichment scores at ATAC-seq modules (ATAC1-5), and proportion of elements in each module containing motif. Note the motif for Hes1 was sourced from the full HOCOMOCO TF motif database (see section 2.15.1, p.56 for database details). Nodes highlighted in red include the Hes1 motif, and TF motifs associated with predicted Runx1 targets, differentially expressed the pre-HE to HE or HE to pro-HSC transitions.

ETS motif TFs (Ets1, PU.1) are enriched in ATAC-seq modules ATAC2, ATAC3 and ATAC5, with some specificity differentiating PU.1 and Ets1 HOCOMOCO PWMs resulting in higher enrichment for the PU.1 PWM in ATAC5 (Fig. 3.13A-B).

Motifs for Meis1 and Nfe2 are enriched in ATAC2 peaks, suggesting that they interact with the endothelial program (Fig. 3.13A-B). Meis1 has been reported to promote EC development in zebrafish (Minehata *et al.*, 2008), and *Meis1* knockout studies result in abnormal vasculature (Hisa *et al.*, 2004). Pbx1 motifs are enriched in both ATAC3 and ATAC4 elements (Fig. 3.13A-B), and Pbx1 is known to mediate HSC quiescence, possibly through activation of Tgf-β related genes (Ficara *et al.*, 2008), and as such may regulate Tgf-β activity in IAC populations. Motifs for Ikzf1, Mef2c and Gfi1 are biased towards ATAC3 (Fig. 3.13A-B), suggesting these factors regulate the transiently accessible module. Mef2c is known for its roles in megakaryopoiesis and B-cell development (Gekas *et al.*, 2009), and Ikzf1 is a TF critically required for B-cell development (Nichogiannopoulou *et al.*, 1999), though their role in EHT is unclear.

3.6.1 A Runx1 and Ikzf1 driven FFL regulates Notch pathway genes

An interesting aspect of Runx1 activity in the EHT GRN model is that Runx1 targets in the RNA3 expression module are anti-correlated with *Runx1* expression (Appendix A.13), suggesting that Runx1 represses genes in the transient network. I next explored the capacity for Runx1, in cooperation with other TFs, to repress transiently expressed genes. In section 3.2.1, I found the transient expression and accessibility modules (RNA3 and ATAC3) to be enriched for Notch signalling by GO analysis (Fig. 3.3C-D). Interestingly, motifs for the Notch target gene, *Hes1*, are also highly enriched at transient elements (ATAC3, Fig. 3.13B). This suggests that Notch signalling is a prominent regulator of the transient EHT network. I screened all TFs enriched in ATAC3 for possible regulation of the Notch pathway, by quantifying the number of GRN predicted targets at Notch signalling related genes (GO:0007219) (Fig. 3.14A). The TFs that were found to most frequently target the Notch pathway within the GRN model were Ikzf1, PU.1 and Ets1. Previous study identified Ikzf1 as a repressor of Notch signalling in T-cell development (Kleinmann *et al.*, 2008), and this analysis suggests that this activity is present in EHT as

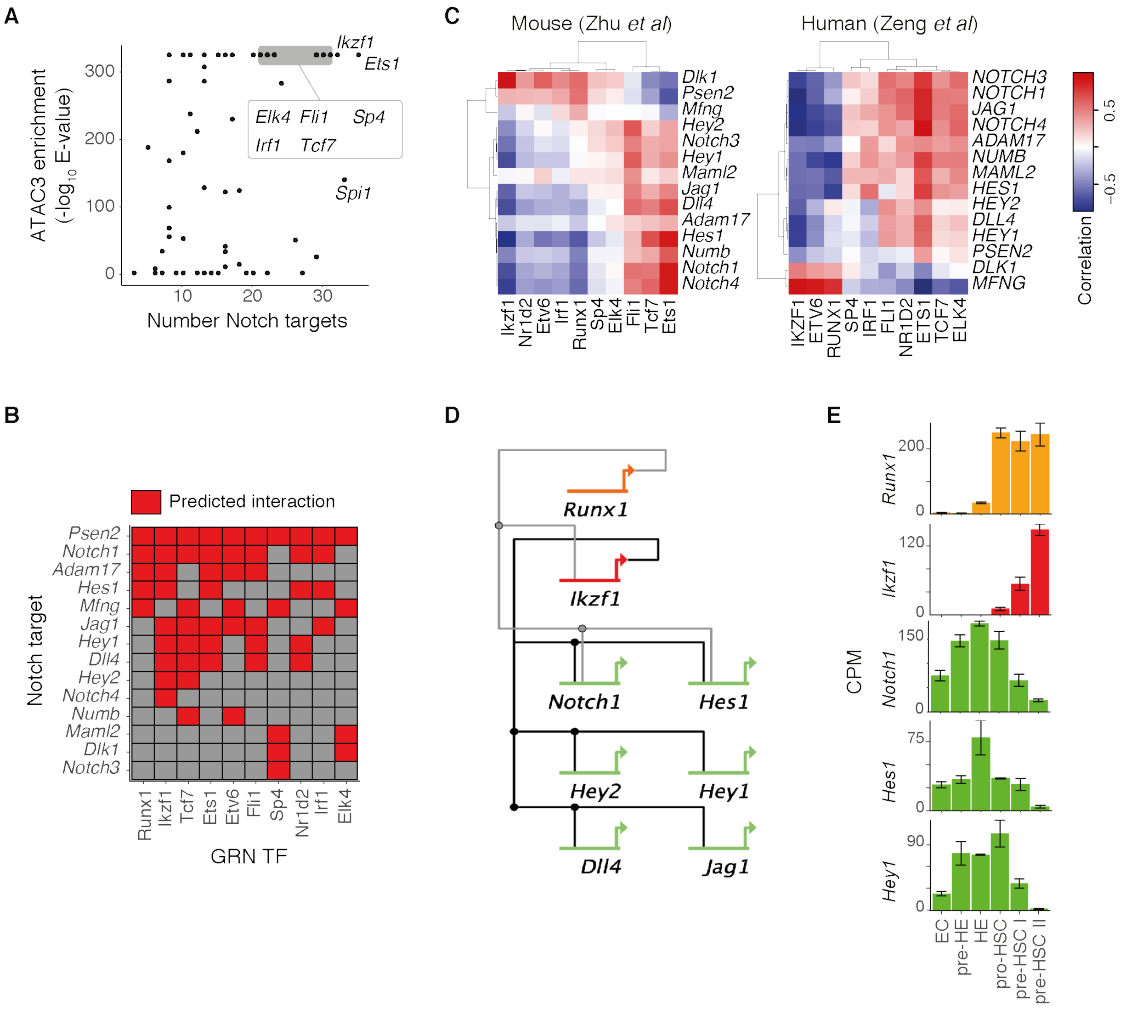


Figure 3.14: Runx1 and Ikzf1 cooperate to repress the Notch pathway through a Runx1:Ikzf1:*Hes1* FFL. (A) Relationship between enrichment of TF motifs at ATAC3 elements, and the number of GRN predicted TF targets in the Notch pathway (GO:0007219). **(B)** TF interaction matrix showing predicted regulation of TFs at Notch pathway targets. TF regulators are shown as columns, and predicted Notch pathway targets as rows. **(C)** Correlation heatmaps from scRNA-seq EHT data. **(D)** Circuit illustrating predicted regulation of Notch genes by Runx1 (light grey lines) and Ikzf1 (dark grey lines). **(E)** Gene expression (CPM) for *Runx1*, *Ikzf1*, and select Notch genes. Error bars represent standard error of the mean; $n = 3$; $n = 4$ for pre-HE. *Raw scRNA-seq data from Zhu* et al.*, 2020 and Zeng* et al.*, 2019, reanalysed by me.*

well. Ikzf1 is of particular interest, as it is predicted to be regulated by Runx1 (section 3.4.3, Fig. 3.10), is found to co-interact with Runx1 (section 3.5, Fig. 3.11), and is a highly central TF within the RNA5 gene expression module (section 3.3.1, Fig. 3.7), suggesting it is an active driver of late-EHT populations. IKZF motifs are significantly enriched at all ATAC-seq modules, but most enriched at

transiently accessible (ATAC3) sites (Fig. 3.13B). As Ikzf1 can deactivate targets through recruitment of HDAC (Marke *et al.*, 2018), and its motif is highly enriched at transiently accessible (ATAC3) sites, this positions it as a potential repressor of the transient EHT network. More specifically, it may repress Notch pathway genes, which is a known behaviour of Ikzf1 in T-cell development (Kleinmann *et al.*, 2008), but is an as yet unexplored interaction in EHT.

As *Ikzf1* is a predicted target of Runx1, and Ikzf1 co-interacts with Runx1, it is possible that these factors cooperate at Notch genes in a FFL circuit. While Ikzf1 is predicted to regulate many Notch genes, Runx1 is predicted to regulate relatively few, such as *Notch1* and *Hes1* (Fig. 3.14B). Ikzf1 also regulates these loci, and additionally targets *Hey1*, *Hey2*, *Jag1* and *Dll4*. I used the mouse and human scRNA-seq EHT data from Zhu *et al.*, 2020 and Zeng *et al.*, 2019, respectively, to compare expression profiles (Fig. 3.14C). *Runx1* expression was strongly anti-correlated with *Hes1* and *Notch1*, but not *Hey1* or *Hey2* (Fig. 3.14C), in line with its predicted regulation. *Ikzf1* was instead anti-correlated with each of its predicted targets. Together, this suggests a key role for Ikzf1 in repressing Notch signalling, and may cooperate with Runx1 at a subset of Notch pathway genes. GRN predicted interactions downstream of Runx1 and Ikzf1 suggest a FFL circuit, where Runx1 activates *Ikzf1*, and both TFs cooperate to repress *Notch1* and *Hes1* (Fig. 3.14D). Additionally, a TF cascade is predicted where Ikzf1 regulates several Notch genes independently of Runx1. Based on expression data, Runx1 does not appear to be sufficient to fully repress *Notch1* or *Hes1* alone, and full inactivation of these genes does not occur until *Ikzf1* is upregulated in pre-HSC I and pre-HSC II populations (Fig. 3.14E). This is in line with the hypothesis that Runx1 and Ikzf1 can cooperate with each other, and may synergise to maximise repression of the Notch pathway.

To functionally validate the existence of these FFL circuits (Fig. 3.14D), it is first important to explore the Runx1 driven regulation of *Ikzf1*. The EHT GRN predicted Runx1 to regulate the -28, -112, and -121 kb *Ikzf1* enhancers upstream of the promoter. The -121 and -28 elements were characterised in Alomairi *et al.*,

2020, where the -121 element was found to be a strong enhancer in $CD4^+CD8^+$ T-cells. The enhancer potential, and relevance of these enhancer elements to EHT, is as yet unclear, but may act to drive Runx1 and Ikzf1 FFLs. To validate that these have enhancer features, I reanalysed published Runx1 and H3K27ac ChIP-seq data in 416B cells (CD34+ myeloid cell line) (Schütte *et al.*, 2016), and capture-C analysis performed by D. Owens (de Bruijn lab). These data found the -28, -112 and -121 elements were bound by Runx1 and marked by H3K27ac, and to physically interact with the *Ikzf1* promoter (Fig. 3.15A). Runx1 was functionally validated to regulate *Ikzf1* by L. Greder (de Bruijn lab), who showed that *Ikzf1* expression was

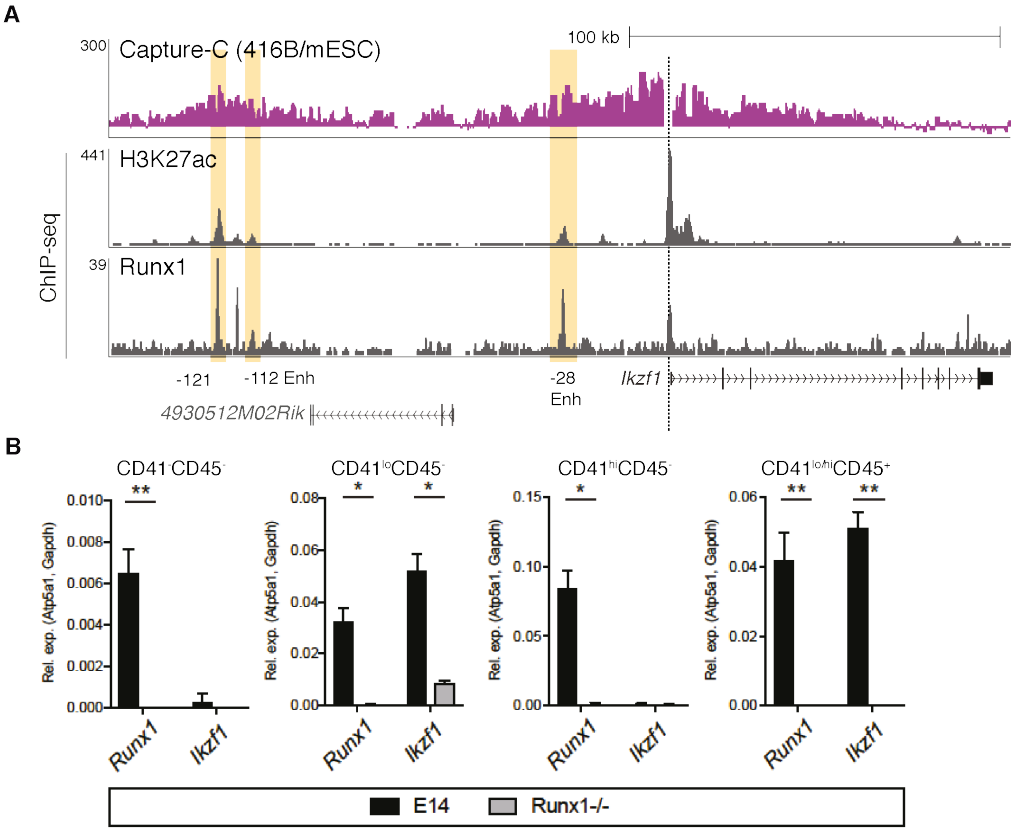


Figure 3.15: Validation of Runx1:*Ikzf1* interaction. (A) Capture-C tracks over the *Ikzf1* locus (top), using the *Ikzf1* promoter as the capture point (dashed line). Reads normalised as 416B signal (active *Runx1*) over mESC signal (inactive *Runx1*). Runx1 and H3K27ac ChIP-seq from 416B cells (bottom). GRN enhancer elements annotated. **(B)** qRT-PCR assaying *Runx1* and *Ikzf1* expression at day 7 of differentiation of $Runx1^{+/+}$ and $Runx1^{-/-}$ mESC lines ($n = 3$). Expression normalised to *Atp5a1* and *Gapdh*. ns: not significant, * $P < 0.05$, ** $P < 0.01$. *Capture-C (A) performed and analysed by D. Owens, and qRT-PCR (B) performed by L. Greder. ChIP-seq data (A) sourced from Schütte* et al.*, 2016 and analysed by me.*

reduced in *Runx1*$^{-/-}$ cells (Fig. 3.15B). Further work from L. Greder (de Bruijn lab, unpublished) has functionally tested the Runx1:Ikzf1:*Hes1* FFL circuit in mESC differentiation cultures (Appendix A.14). This experiment used *Ikzf1*$^{-/-}$ and *Ikzf1*$^{+/+}$ mESCs, and used a *Runx1-ERt2* fusion to control Runx1 nuclear translocation. Ikzf1 and Runx1 activity is referred to here as active (Ikzf1+/Runx1+) or inactive (Ikzf1-/Runx1-). It was found that in Ikzf1-/Runx1+ conditions, Runx1 appears to promote *Hes1* expression over Ikzf1-/Runx1- conditions (Appendix A.14), contrary to the GRN-predicted activity of Runx1 at this locus. In Ikzf1+/Runx1- conditions, *Hes1* is instead repressed compared with Ikzf1-/Runx1- conditions, in line with the predicted repressive activity of Ikzf1. Interestingly, *Hes1* is maximally repressed only in Ikzf1+/Runx1+ conditions, suggesting that both Runx1 and Ikzf1 are required for full downregulation of the gene. These analyses build evidence for a functional Runx1:Ikzf1:*Hes1* FFL circuit that drives repression of *Hes1*, and is a potential mechanism to drive Notch pathway suppression in pre-HSC I and pre-HSC II populations where *Ikzf1* is expressed.

3.7 Maf and Lef1 cooperation may drive cytoskeletal remodelling and EMT processes

The data thus far showcases the utility of the analytical strategy for predicting novel cooperative regulation driving key processes. I next aimed to explore potential TF cooperation involving factors other than Runx1. During EHT, the morphological changes and budding of HE is an important process for the formation and maturation of IACs. Among TF co-interaction clusters, TFs in CoInt-4 were transiently expressed (RNA3) and enriched for blood vessel morphogenesis and EMT (Fig. 3.11B-C), reflecting the EHT budding process. TFs in CoInt-4 also have an overall greater proportion of targets within actin cytoskeletal remodelling and EMT pathways than other co-interaction clusters (Fig. 3.16A). Of these, the Maf family of TFs is predicted to regulate a high number of both actin cytoskeletal remodelling and EMT genes. Maf and Maff are bZIP TFs that share a highly similar

DNA binding motif. Maf forms homodimers and contains a potent transactivating domain, while Maff lacks a transactivating domain but regulates gene targets through heterodimerising with other bZIP TFs (Kataoka, 2007; Katsuoka *et al.*, 2016). As noted earlier, MAF motifs are found in the +3 and +23 *Runx1* enhancers (Fig. 3.8B), suggesting they may play a role in the activity of these enhancers.

There are several genes that significantly co-interact with Maf, but relatively few of these prominently interact with actin or EMT related genes (Fig. 3.16B). Lef1 is a CoInt-3 TF that interacts highly with both pathways and is among the most central regulators of the transient expression module (Fig. 3.7A). Of actin genes, both *Maf* and *Lef1* gene expression is highly correlated with *Marcks* and *Nckap1* expression, and moderately correlated with *Cdc42ep1* (Fig. 3.16C). Marcks is a membrane-bound or cytoplasmic protein that drives cytoskeletal rearrangement and migration, and has been implicated in EMT (El Amri *et al.*, 2018; Xiang *et al.*, 2019). Cdc42ep1 (also known as Borg5) is associated with Cdc42, a key regulator of actin dynamics, and drives directional migration and angiogenesis (Farrugia *et al.*, 2016; Liu *et al.*, 2014; Watson *et al.*, 2016). Nckap1 (HEM-2) is a component of the WAVE complex, which promotes activity of Arp2/3 to drive actin branching (Chen *et al.*, 2014; Chen *et al.*, 2010). To reinforce this model of Maf and Lef1 regulating actin related genes I used the EHT scRNA-seq data from Zhu *et al.*, 2020 and Zeng *et al.*, 2019. Many of the actin-related genes are correlated with each other, as well as with *Maf* and *Lef1* expression (Fig. 3.16D). Note that these scRNA-seq data consist of E9.5 and E10.5 populations, and so captures the end of EHT. Single cell pseudotime analysis shows *Maf*, *Lef1*, and actin genes to decrease in expression over time, in line with bulk RNA-seq expression patterns from E9.5 HE to E10.5 pre-HSC II (Fig. 3.16E-F). This analysis supports a model whereby Maf1 and Lef1 cooperate to drive morphological changes, though this has yet to be experimentally validated.

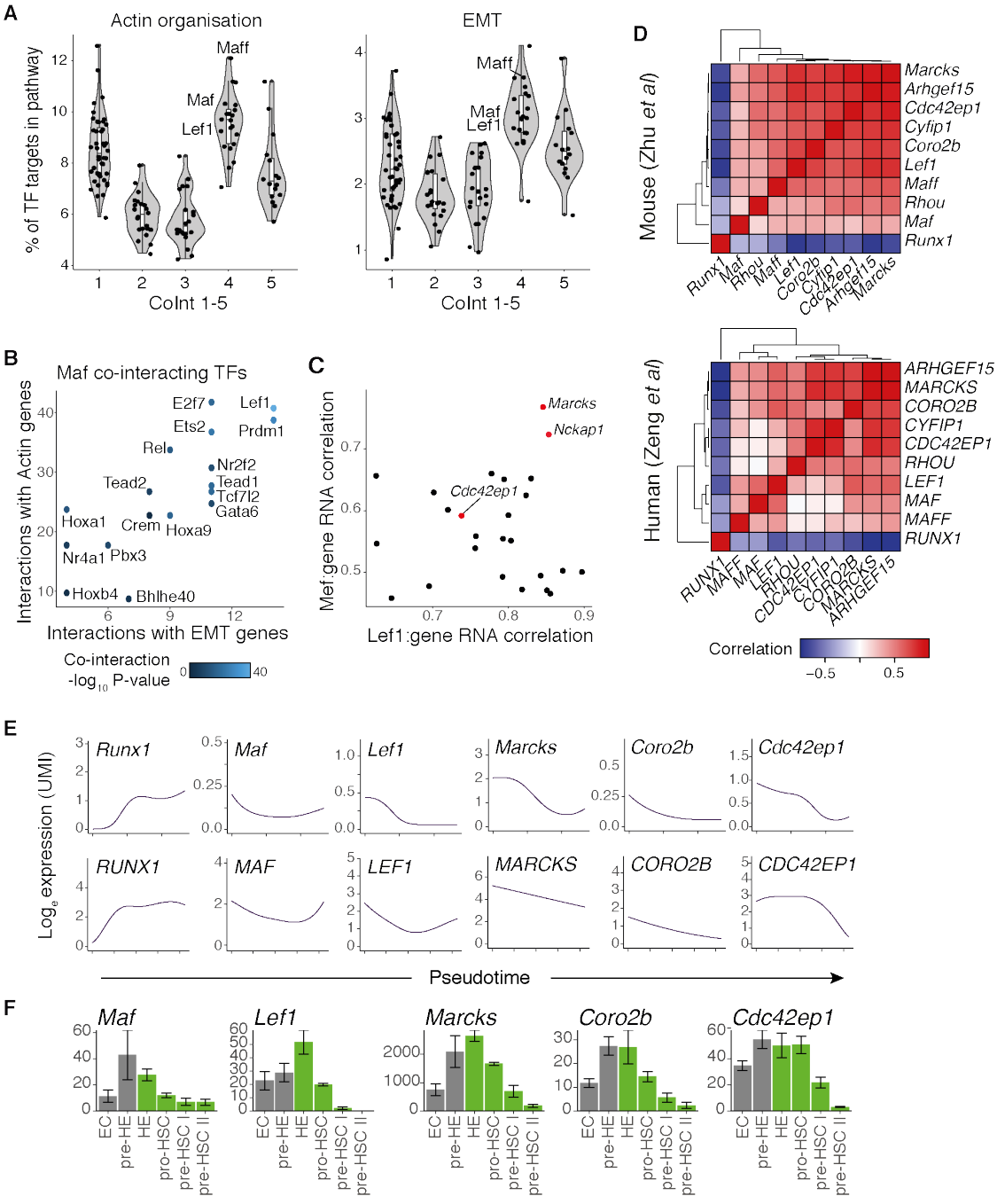


Figure 3.16: The transient co-interaction module is predicted to drive cytoskeletal rearrangement through Maf:Lef1 cooperative action. (A) Proportion of TF targets that regulate actin cytoskeletal organisation (GO:0030036, left) and EMT (GO:0001837, right), across co-interaction clusters. Colour represents RNA expression modules. **(B)** Frequency of TF targets regulating actin cytoskeletal organisation and EMT pathways. TFs shown significantly co-interact with Maf (FDR < 0.05). **(C)** Gene expression correlation of actin cytoskeletal organisation genes, and *Maf* or *Lef1*. Genes of interest highlighted in red. **(D)** Correlation heatmaps from scRNA-seq EHT data. *Legend continued on next page.*

Figure 3.16: *Legend continued from previous page.* **(E)** scRNA-seq expression of *Maf*, *Lef1*, and key actin related genes for mouse (top) and human (bottom) over pseudotime. Lines indicate smooth fitted values calculated by TradeSeq. **(F)** Gene expression (CPM) for *Maf*, *Lef1*, and key actin related genes in bulk RNA-seq data. Error bars represent standard error of the mean; $n = 3$; $n = 4$ for pre-HE. *Raw scRNA-seq data from Zhu* et al.*, 2020 and Zeng* et al.*, 2019, reanalysed by me.*

3.8 Conclusions and discussion

In this chapter, I aimed to explore potential mechanisms by which EHT regulation is driven. Through the integration of RNA-seq and ATAC-seq datasets, I have established a GRN model that predicts TF regulation that may drive activity of complex regulatory patterns associated with EHT. Using this GRN model I characterised predicted upstream regulators, and downstream targets, of Runx1. Taking advantage of these predicted interactions, I performed an analysis predicting possible TF cooperation that may drive EHT programs. One of these predicted TF cooperations was a Runx1:Ikzf1 driven FFL that represses Notch pathway genes. This FFL may contribute to deactivation of the transiently regulated EHT program, as the Hes1 motif underlies many transiently accessible enhancers (Fig. 3.13B).

While several studies have established GRN models of EHT (Baron *et al.*, 2018; Bergiers *et al.*, 2018; Gao *et al.*, 2020; Zeng *et al.*, 2019; Zhu *et al.*, 2020), they do not incorporate early EHT populations in E8.5 embryos. The initiation of HE identity can be considered to occur at E8.5, with activation the 23GFP transgene in dorsal aorta endothelium (Swiers *et al.*, 2013a), referred to as pre-HE. 23GFP expression in pre-HE precedes endogenous *Runx1* transcription, and was shown to receive pro-haemogenic signals prior to *Runx1* activation. Analysis of RNA-seq data found *Cd109* to be highly expressed in pre-HE and HE cells, and inactive in 23GFP- endothelium (section 3.2.2). The role of CD109 in EHT has not been studied, though it is known to block TGF-β signalling (Mii *et al.*, 2019), and therefore may act to modulate this pathway in pre-HE and HE. The role of TGF-β signalling in EHT is unclear, as it has been shown to both inhibit (Vargel *et al.*, 2016), and promote EHT (Monteiro *et al.*, 2016). Importantly, Monteiro *et al.*, 2016 used morpholinos in zebrafish to perturb *Tgf-β* and *tgfbR2*, with no population-specific targeting, while Vargel *et al.*, 2016 treated with Tgf-β ligand or inhibitor in mESC haematopoietic differentiations. Both studies show that either addition of TGF-β (Vargel *et al.*, 2016), or removal of Tgf-β (Monteiro *et al.*, 2016), result in impaired EHT. These studies perturb Tgf-β to either low, or higher than normal levels, which

may indicate a specific dose-dependent requirement that Cd109 may function to adjust in HE. Further study is required to determine the role of CD109 in EHT, and how TGF-β signalling is affected by its expression.

The GRN model was built through integrating multiple experimental datasets (section 3.3), using a methodology with specific advantages and disadvantages. The construction of this GRN relied upon robust identification of regulatory elements, and reliable association of these elements to promoters. Due to the scarcity of HE cells in embryos, ChIP-seq data was unavailable and so enhancer histone marks (H3K27ac and H3K4me1, Creyghton *et al.*, 2010; Heintzman *et al.*, 2009; Heintzman *et al.*, 2007) could not be used. Additionally, while DAEs were annotated to the nearest DEG, regulatory elements are known to skip promoters to act on distal genes (Chepelev *et al.*, 2012). These caveats were compensated for by incorporating only DEGs and DAEs, so that dynamic genes and accessible elements are incorporated, and E-P interactions were filtered for correlation between RNA-seq and ATAC-seq counts (as used in Hariprakash *et al.*, 2019; Sheffield *et al.*, 2013). As such, E-P interactions were enriched for elements that are likely to regulate gene expression, and as both positive and negative correlated E-P interactions were retained, these elements represent both enhancers and silencers. TF binding motif analysis was used as the basis to infer upstream regulators, which was further reinforced using transcription correlation between the TF regulator and predicted target gene. An additional caveat is TF motifs do not perfectly predict DNA binding. TF binding is multifaceted, and for example is influenced by chromatin accessibility, co-factors that alter base-pair preference, and the formation of cooperative TF complexes that adjust binding affinity (Morgunova *et al.*, 2017; Spitz *et al.*, 2012). The core HOCOMOCO TF motif database was used to establish GRN interactions, as opposed to the full database which contains lower confidence TF PWMs, to improve the reliability of predictions. For motif enrichment analyses (as in Fig. 3.13), the full database was used, containing lower confidence TF PWMs, such as for Hes1. While the approach used in section 3.3 established a computationally

robust GRN model of the overall EHT process, these caveats are important to consider and findings must be experimentally validated.

Gene expression and chromatin accessibility over EHT change dynamically, and can be clustered into distinct modules (RNA1-5 and ATAC1-5, section 3.2.1, Fig. 3.3). These modules represent DEGs and DAEs biased towards different cell populations, and can be categorised as genes and elements that switch into active or inactive states (RNA2/ATAC2, RNA4/ATAC4, and RNA5/ATAC5), and those that are transiently activated or deactivated (RNA1/ATAC1 and RNA3/ATAC3). The transiently active module (RNA3) was found to be associated with EMT and actin cytoskeletal rearrangement processes, as well as Notch and BMP signalling pathways. Conversely, modules that switch into an active state (RNA4 and RNA5) were associated with general haematopoiesis. By integrating these modules with the GRN, key drivers of each module could be predicted (section 3.3.1, Fig. 3.7). Degree centrality was used to predict key drivers within each module, and is known to correlate with gene essentiality in other systems (Hahn *et al.*, 2005; Jeong *et al.*, 2001; Koschützki *et al.*, 2008). This highlighted many known EHT regulators, particularly in the RNA4 module, including Runx1, Gfi1, and PU.1. Central TFs in the transiently expressed module (RNA3) included factors such as Fli1, Nfib, Tead1, and Gata3/6. Gata3 has recently been found to have an important role in aortic endothelium for EHT (Zaidan *et al.*, 2022), which in combination with its transient expression pattern, may suggest a role in HE regulation. RNA5 TFs are expressed primarily in pre-HSC populations, and the top-most central of these factors include Ikzf1, which has not yet been extensively studied in EHT.

As Runx1 is essential for EHT, I explored both upstream and downstream regulators (section 3.4.2 and 3.4.3) of the TF. As the pre-HE population is characterised by 23GFP expression, it implies that factors that can regulate enhancer activity are present at this stage. There are multiple TFs with motifs found in the +23 enhancer, that are also upregulated in pre-HE over ECs, including Maf, Elk4, Klf6, Prdm16, Smad3, Sp4, and Tcf4, with *Maf* expression being particularly

biased towards pre-HE. At transitions where *Runx1* is upregulated (pre-HE to HE, and HE to pro-HSC), we find several TFs concurrently upregulated in these transitions, with motifs at *Runx1* enhancers. For example, NFI motifs underly -59 and +24 enhancers, and Nfib is a highly central RNA3 TF. Data from the de Bruijn lab (unpublished) found the +24 element to have little enhancer activity on its own, but boosted the activity of the +23 enhancer, so Nfib may act to support +23 enhancer activity. Additionally, CEBP motifs were found specifically in the +110 enhancer, and AP-1 motifs specifically in the +204 enhancer, highlighting the potential differences between regulatory elements.

One aspect of gene regulation that remains challenging to study, is the cooperative action of TFs to regulate genes. TFs can often act in complexes, resulting in changes in their activity or shifting DNA binding sequence preference. Runx1 is known to cooperate with multiple TFs (Goetz *et al.*, 2000; Hu *et al.*, 2011; Kim *et al.*, 1999b; Lichtinger *et al.*, 2012; Wang *et al.*, 2011; Zaidi *et al.*, 2002). I explored TFs that frequently co-interact with Runx1 at enhancers and promoters (section 3.5), suggesting that Runx1 may cooperate with these factors to co-regulate genes. Among the TFs most frequently co-interacting with Runx1 was Ikzf1, a highly central regulator in the RNA5 expression module. *Ikzf1* was also found to be a target of Runx1 (section 3.4.3, Fig. 3.10). While *Ikzf1* was expressed in pre-HSC populations, IKZF1 motifs were found primarily in transiently accessible elements (ATAC3, section 3.6.1, Fig. 3.13). The GRN analysis predicted Runx1 and Ikzf1 to repress the transiently active Notch pathway (RNA3, ATAC3) in a FFL configuration, and experiments performed by L. Greder (de Bruijn lab, unpublished) supported these factors to repress *Hes1* expression (Fig. 3.17, Appendix A.14). Interestingly, these experiments found that in the absence of Ikzf1, Runx1 appeared to upregulate *Hes1* expression. In the presence of Ikzf1, Runx1 instead represses *Hes1* expression. This could suggest that the activity of Runx1 changes, depending on the presence of Ikzf1. One possible explanation for this is that Runx1 and Ikzf1 may physically interact, resulting in a change in Runx1 activity.

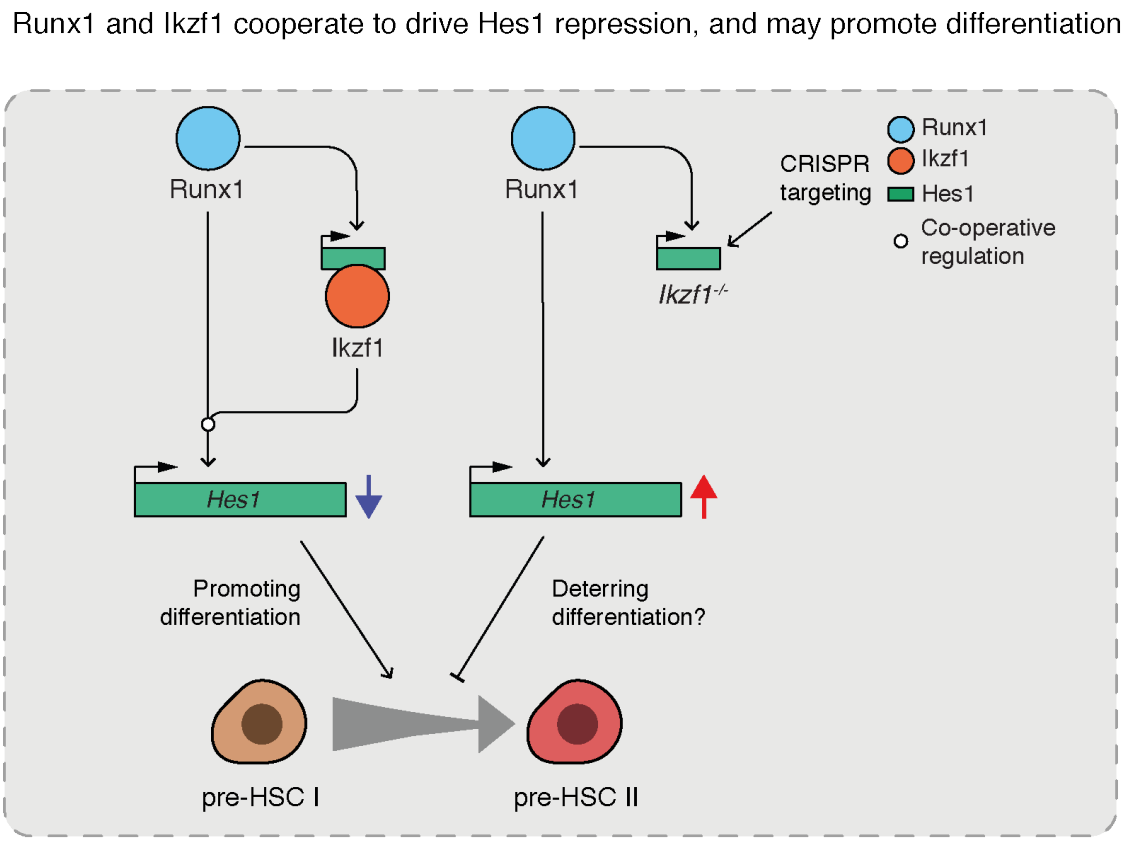


Figure 3.17: Summary model of Runx1 and Ikzf1 cooperation to drive *Hes1* repression. A model of a Runx1 and Ikzf1 driven FFL, predicted to drive *Hes1* repression. Runx1 and Ikzf1 appear to cooperate at the *Hes1* locus. Where Runx1 and Ikzf1 are predicted to cooperate and drive repression, in *Ikzf1*$^{-/-}$ conditions Runx1 may instead upregulate *Hes1*. As repression of the Notch pathway is required for the pre-HSC I to pre-HSC II transition (Souilhol *et al.*, 2016), this may be an important circuit for the progression of late-EHT maturation.

A previous study has found human IKZF1 to physically interact with RUNX1 protein (Zhou *et al.*, 2019), supporting this hypothesis. An alternative explanation is that Ikzf1 activity takes precedence, and as Runx1 drives *Ikzf1* expression, there may be higher levels of Ikzf1 in Ikzf1+/Runx1+ conditions resulting in greater *Hes1* downregulation. These functional analyses support a mechanism by which Runx1 and Ikzf1 repress Notch pathway genes in pre-HSC I and pre-HSC II, though further study is necessary to elucidate the molecular mechanism behind Runx1 and Ikzf1 cooperation. Importantly, repression of the Notch pathway is a required step for the progression of pre-HSC I to pre-HSC II cells (Souilhol *et al.*, 2016), placing this Runx1:Ikzf1:*Hes1* FFL as a potentially important circuit for driving

differentiation of late EHT populations (Fig. 3.17).

The analytical strategy for exploring TF co-interaction was also applied to study the regulation of EMT and actin cytoskeletal organisation pathways, in order to investigate the processes underlying the morphological changes that occur during EHT (section 3.7). This preliminary analysis identified Maf and Lef1 as two TFs that were predicted to regulate many genes in these pathways, and these were found to frequently co-interact at the same accessible elements. These factors were also highly correlated with several actin cytoskeleton genes, particularly *Marcks*. The Marcks protein drives cytoskeletal rearrangement, resulting in cell migration and EMT (El Amri *et al.*, 2018; Xiang *et al.*, 2019), placing it as a candidate driver of EHT morphological changes. Maf and Lef1 were also predicted to regulate Cdc42ep1 (Borg5), which associates with an important driver of cell migration, Cdc42 (Farrugia *et al.*, 2016; Liu *et al.*, 2014; Watson *et al.*, 2016). This places both Maf and Lef1 as possible drivers of cytoskeletal machinery, and may contribute to the budding process during EHT. Further study is required to validate these interactions, and determine whether Maf1 and Lef1 functionally cooperate, and whether their activity is required for EHT.

4

MLL-AF4 and RUNX1 driven regulatory networks in *MLL*r leukaemia

Contents

4.1 Introduction and aims 105
4.2 Generating and characterising a GRN model of MLL-AF4 driven leukaemia 108
4.3 Integrating patient expression data highlights regulatory modules, and a set of "core" MLL-FP TFs 113
4.3.1 Regulatory modules of the GRN 115
4.3.2 High centrality nodes represent a core MLL-FP program 116
4.4 Validating the importance of central regulatory hubs . 119
4.4.1 Essentiality screens validate the importance of several central TFs 119
4.4.2 Stress is a better indicator of essentiality than degree centrality 121
4.5 MLL-AF4 and RUNX1 cooperate to regulate gene expression 123
4.5.1 Generating, and characterising a RUNX1 driven GRN model of leukaemia 123
4.5.2 RUNX1 and MLL-AF4 GRNs intersect at leukaemogenic processes 125
4.6 MLL-AF4 and RUNX1 direct both FFL and TF cascade regulation 126
4.6.1 Identifying FFL and cascade regulatory circuits 126
4.6.2 Correlation between MLL-AF4 and RUNX1 ChIP-seq binding profiles 130
4.7 TF cooperation at MLL-AF4 targets drives expression and promotes leukaemogenesis 133
4.8 MLL-AF4 and cooperative TFs co-opt and drive normal fetal circuits. 136
4.9 Conclusions and discussion 145

4.1 Introduction and aims

In chapter 3 I established a GRN model of mouse EHT. This is a normal haematopoietic process, describing the differentiation of aortic endothelium towards HSCs. I focused the analysis of this model on the activity of Runx1, and I used an analytical approach to predict possible TF cooperation. For this chapter, I asked how abnormal haematopoiesis may be regulated. MLL-AF4 leukaemias are considered a transcriptional disease, as the translocation event can be sufficient on its own for leukaemogenesis, with few or no cooperative mutations (Andersson *et al.*, 2015; Bardini *et al.*, 2011; Bardini *et al.*, 2010; The Cancer Genome Atlas Research Network, 2013). As described in section 1.4.1, MLL-AF4 binds to gene promoters and drives transcription through the recruitment of a large transcription elongation complex (Ballabio *et al.*, 2012; Slany, 2020). This makes *MLL*r leukaemias particularly suited for analysis using GRN methods, which rely on transcriptional data. Additionally, the MLL-AF4 fusion protein maintains overexpressed *RUNX1* expression (Wilkinson *et al.*, 2013), which allows for the study of dysregulated RUNX1 activity in a leukaemic network context.

As the MLL-AF4 complex directly promotes transcription (Ballabio *et al.*, 2012; Biswas *et al.*, 2011; Kerry *et al.*, 2017; Lin *et al.*, 2010; Mueller *et al.*, 2007; Okuda *et al.*, 2015; Yokoyama *et al.*, 2010) and is the main driver of leukaemogenesis, it is expected that *MLL-AF4* KD would lead to downregulation of gene expression, with affected genes bound by MLL-AF4 protein. Previous analysis by R. Thorne (Milne lab) of nascent RNA-seq following *MLL-AF4* KD (Godfrey *et al.*, 2017) in the SEM cell line (B-ALL MLL-AF4, Greil *et al.*, 1994) established that this is not true. Only 28% of *MLL-AF4* KD sensitive genes were bound by MLL-AF4, and both up- and downregulation was observed (Fig. 4.1A, Harman *et al.*, 2021). To explain this discrepancy, I hypothesised these non-bound genes are indirectly regulated via intermediate TFs that are overactivated as a consequence of MLL-AF4 activity (Fig. 4.1B). As TFs often do not act in isolation (Morgunova *et al.*, 2017; Spitz *et al.*, 2012), it is important to address the possibility that MLL-AF4 may function in

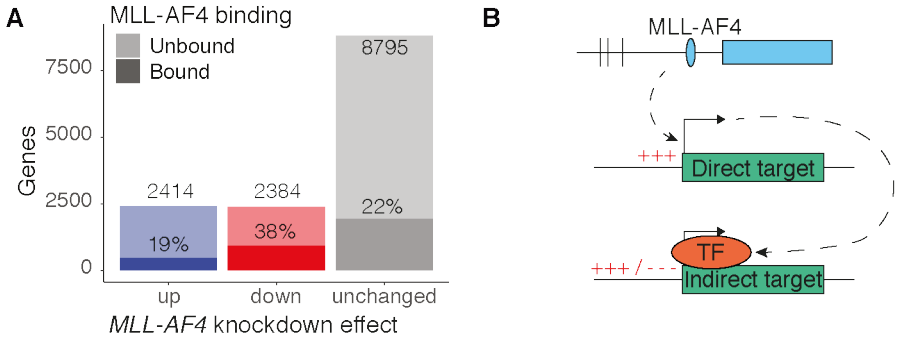


Figure 4.1: MLL-AF4 regulation cannot be explained by direct binding alone. (A) Number of DEGs after 96 h *MLL-AF4* KD in SEM cells ($n = 3$). Shaded bar represents proportion of DEGs bound by MLL-AF4 (ChIP-seq). **(B)** Schematic illustrating hypothesis, where MLL-AF4 regulates TFs, which then regulate additional targets, resulting in indirect downstream regulation. *Analysis in A first performed by R. Thorne. Adapted from Harman* et al.*, 2021.*

combination with other TFs to regulate gene targets, and thus understand the full regulatory scope of MLL-AF4. As such, a GRN model of MLL-AF4 leukaemia was established to decode complex combinatorial signals and organise direct and indirect regulatory logic, by isolating common network motifs such as FFLs and TF cascades.

Aims

1. To determine if MLL-AF4 controls a wider transcriptional network through TF activation.

2. To determine whether the MLL-AF4 GRN can be used to identify novel biology of MLL-AF4 leukaemias.

3. To investigate the TF cooperativity between MLL-AF4, RUNX1, and its target TFs.

By constructing a GRN model of MLL-AF4 leukaemia, the regulatory logic downstream of MLL-AF4 can be explored. This allows us to address the hypothesis that MLL-AF4 regulates targets indirectly via intermediate TFs. It is important to determine whether the GRN can describe the biology of leukaemia, and its relevance to real-world patient samples. As in chapter 3, a dissection of how MLL-AF4 and RUNX1 cooperate with other TFs will explore into a less studied area

of gene regulation, and could offer insight into the role of intermediate TFs in MLL-AF4 leukaemia.

4.2 Generating and characterising a GRN model of MLL-AF4 driven leukaemia

To determine whether MLL-AF4 was able to control the expression of a wider transcriptional network by activating TFs, I sought to integrate nascent RNA-seq and ChIP-seq datasets generated in SEM cells (B-ALL MLL-AF4) to establish a GRN model. MLL-AF4 DNA binding is not well predicted through specific motifs. The part of MLL-AF4 that contacts DNA is the CxxC domain which confers binding preference for uCpGs (Birke *et al.*, 2002). However, MLL-AF4 only binds to a subset uCpGs and it is unknown what guides this preference (Kerry *et al.*, 2017). As such, MLL-AF4 ChIP-seq (Kerry *et al.*, 2017) was used as evidence for direct binding, while nascent RNA-seq following 96 hours' *MLL-AF4* KD (Kerry *et al.*, 2017) was used to identify the regulatory scope of MLL-AF4, including indirect targets.

To establish a robust GRN, it was important that the MLL-AF4 ChIP-seq in the SEM cell line model be of sufficient quality for reliable peak calling, and representative of MLL-AF4 binding profiles in patients. To test this, I compared MLL-AF4 binding in the SEM cell line with bound genes in two MLL-AF4 ALL patient samples. The majority of peaks were common to all three samples (70.4%) with only 9.5% of peaks unique to SEM cells (Fig. 4.2A), and ChIP-seq signal showed high correlation at gene promoters (Fig. 4.2B). Some differences in MLL-AF4 bound genes were observed between the patients (13.8% and 6.2% of SEM bound genes only bound in patient #1 or #2, respectively, Fig. 4.2A), however this is mostly due to lack of precision with peak calling where MLL-N ChIP-seq signal is weak and there are biases in input tracks (Fig. 4.2C-D). Little bias is observed in input tracks for SEM MLL-AF4 or RUNX1 ChIP-seq (Fig. 4.2E). From these analyses, the ChIP-seq data in SEM cells is a good representation of MLL-AF4 binding in patient samples.

To establish the MLL-AF4 GRN (methods section 2.15.3, p.57) I integrated *MLL-AF4* KD DEGs, SEM MLL-AF4 ChIP-seq, and an existing TF interaction network

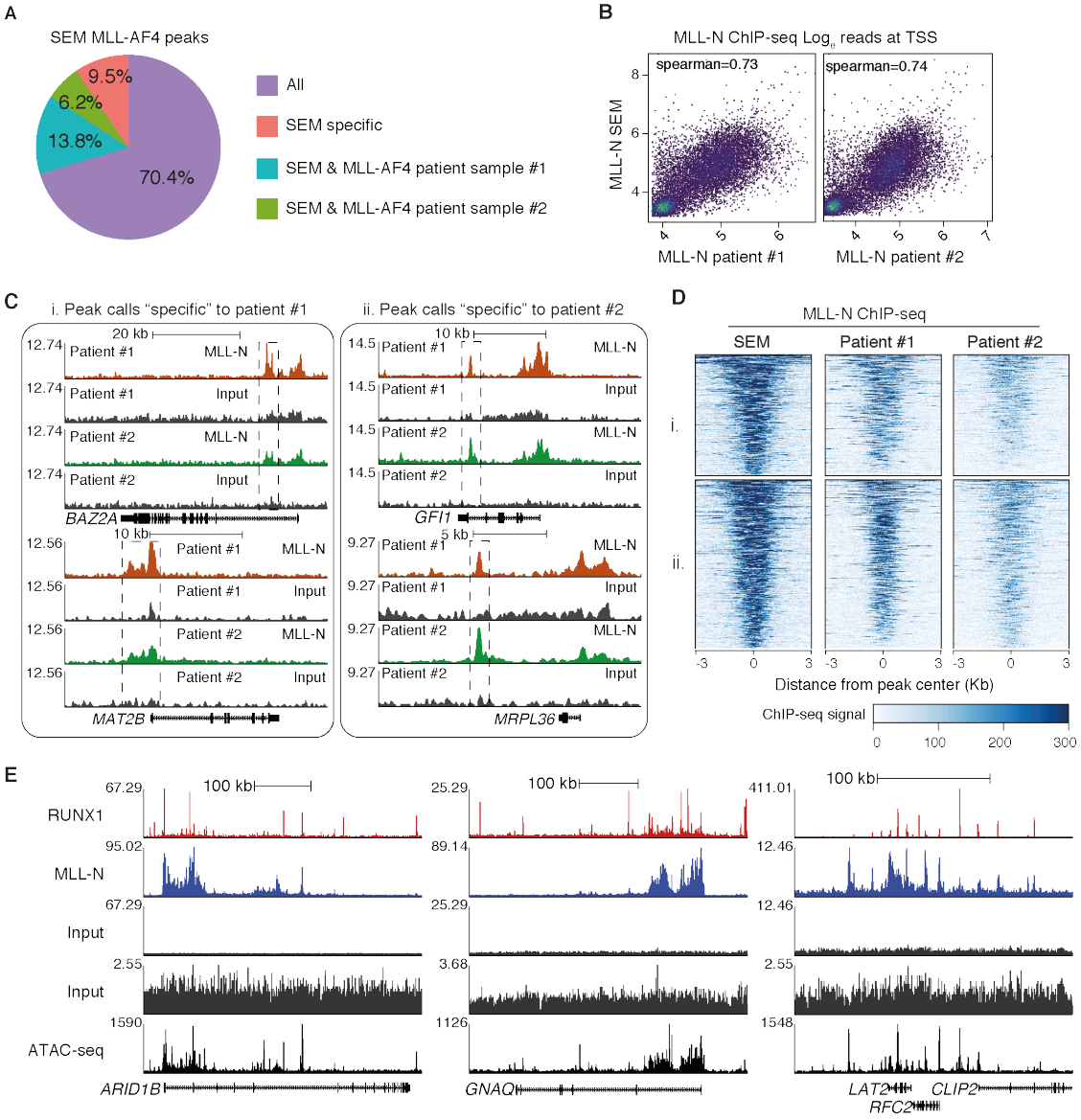


Figure 4.2: ChIP-seq data is of sufficient quality for GRN analysis. (A) Proportion of MLL-AF4 peaks in SEM cells that overlap with MLL-AF4 peaks in patient samples. **(B)** Comparison of Log$_e$ MLL-N reads at gene promoters in SEM cells and patient samples. **(C)** ChIP-seq tracks for MLL-N and input chromatin. Loci are MLL-AF4-bound genes unique to patient #1 (i) or patient #2 (ii), as defined in A. Boxed regions indicate peak calls of interest. **(D)** ChIP-seq heatmaps for MLL-N. Loci are SEM MLL-N peaks with peak calls unique to patient #1 (i) or patient #2 (ii). **(E)** ChIP-seq tracks for MLL-N, RUNX1, and input chromatin, and ATAC-seq from SEM cells. Input tracks are shown scaled to maximum signal, or to the lowest track height in RUNX1 or MLL-N ChIP-seq. RUNX1 and MLL-AF4 samples share input chromatin. *Patient MLL-N ChIP-seq generated by N. Crump, analysed by me. Adapted from Harman* et al., *2021.*

(The FANTOM Consortium and the Riken Omics Science Center, 2009) (Fig. 4.3A, B). This TF interaction network was generated in THP-1 cells (an MLL-AF9 cell line) using deepCAGE sequencing to estimate motif activity. Note that while this underlying network was generated in MLL-AF9 leukaemia we hypothesised that different MLL-FPs would target similar genes, and as such is a sufficient representation of MLL-AF4 biology. The overlap between MLL-AF4 and MLL-AF9 interactions is discussed later in this chapter (See Fig. 4.7). In brief, DEGs from *MLL-AF4* KD were considered to represent the limits of the larger network, and ChIP-seq data was used to connect direct targets of MLL-AF4. The TF interaction network supplemented the ChIP-seq profiles and bridged direct interactions to indirect targets, generating an information rich network of potential regulators.

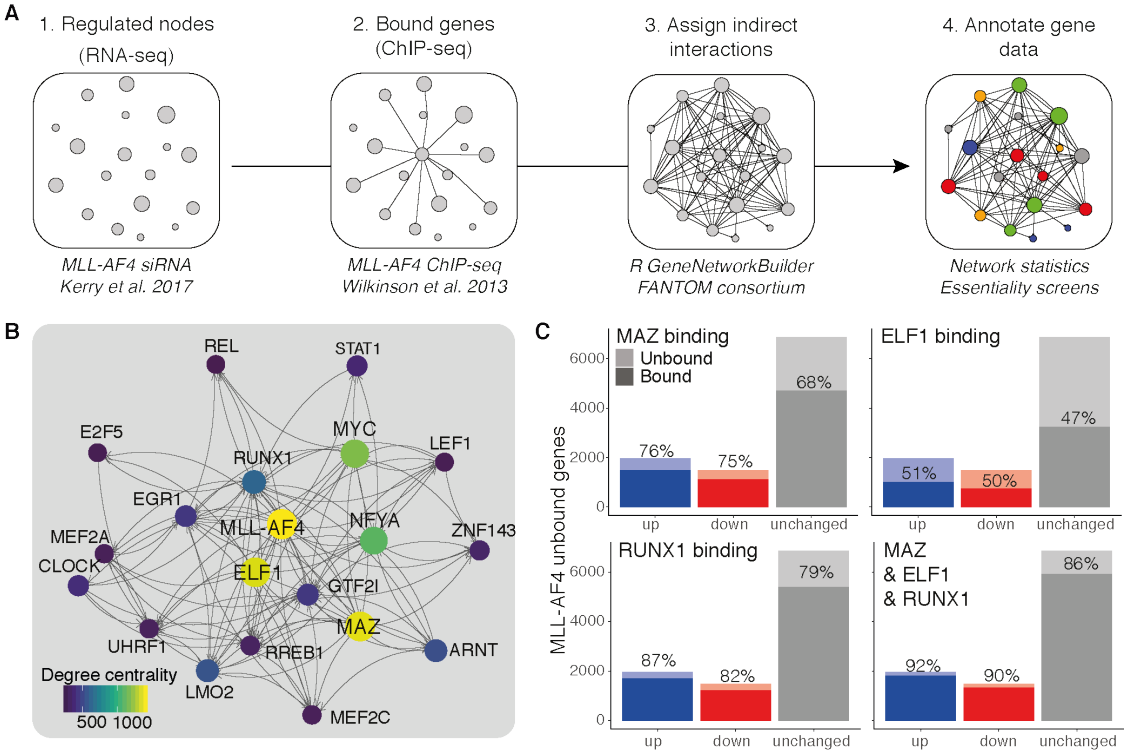


Figure 4.3: Constructing an MLL-AF4 centred GRN model of MLL-AF4 leukaemia. (A) GRN creation workflow using siRNA KD with nascent RNA-seq, ChIP-seq, an underlying TF motif network (The FANTOM Consortium and the Riken Omics Science Center, 2009). **(B)** The top 20 genes of the MLL-AF4 GRN by degree centrality. Lines indicate predicted interaction from protein to gene locus, with arrowheads pointing downstream. **(C)** Number of *MLL-AF4* KD DEGs that are not bound by MLL-AF4, as highlighted in Fig. 4.1A. Shaded bars represent proportion of TF-bound DEGs. *Adapted from Harman* et al., *2021.*

As the central hypothesis of this chapter suggests that indirectly regulated MLL-AF4 targets are affected through activated TFs, MLL-AF4 regulated TFs in the GRN must therefore bind to these indirect targets. Within the GRN a number of highly central TFs were found to be directly regulated by MLL-AF4, including ELF1, MAZ, MYC, NF-YA and RUNX1 (Fig. 4.3C). The Milne lab has previously shown *ELF1* and *RUNX1* to be MLL-AF4 targets, with a role in leukaemia (Godfrey *et al.*, 2019; Wilkinson *et al.*, 2013). Interestingly, Elf1 was also identified as a central regulator for pre-HSCs in the EHT GRN (Fig. 3.7, p.76), suggesting that this role may be co-opted by MLL-AF4 in leukaemia. While MAZ has not previously been implicated in *MLL*r leukaemias, it is known to bind and regulate *MYC* (Komatsu *et al.*, 1997). MYC has a well-studied and critical role in many cancers (Dang, 2012), supporting its place as a central TF here. NF-YA has been recently reported to have a role in *MLL*r leukaemias, in that it may mediate binding of the MLL-FP complex component Menin, and subsequently the MLL-FP complex, to DNA. Inhibition of NF-YA may function synergistically with Menin-MLL1 inhibitors (Soto-Feliciano *et al.*, 2022). Taking ELF1, RUNX1 and MAZ as examples, the binding profiles of these central TFs can account for the majority of MLL-AF4 indirectly regulated targets, and their combined binding profile has the potential to regulate over 90% of MLL-AF4 KD DEGs (Fig. 4.3D), highlighting the utility of this model for interrogating the wider regulatory scope of MLL-AF4.

To validate this GRN model with an alternative MLL-AF4 ALL cell line, I determined how sensitive the GRN nodes are to *MLL-AF4* perturbation in RS4;11 cells. I performed 96 hours' *MLL-AF4* KD in both RS4;11 and SEM cells, and followed this with nascent RNA-seq. The siRNA designed for the MLL-AF4 breakpoint in RS4;11 was much less efficient than the siRNA for SEM cells, despite equivalent treatment conditions (Fig. 4.4A). This reduced efficacy has been observed in other studies (Geng *et al.*, 2012; Thomas *et al.*, 2005). Despite this reduced effectiveness of the siRNA, nascent RNA-seq analysis identified 330 DEGs, and 202 of these overlapped with the SEM MLL-AF4 GRN (Fig. 4.4B). Importantly, even in

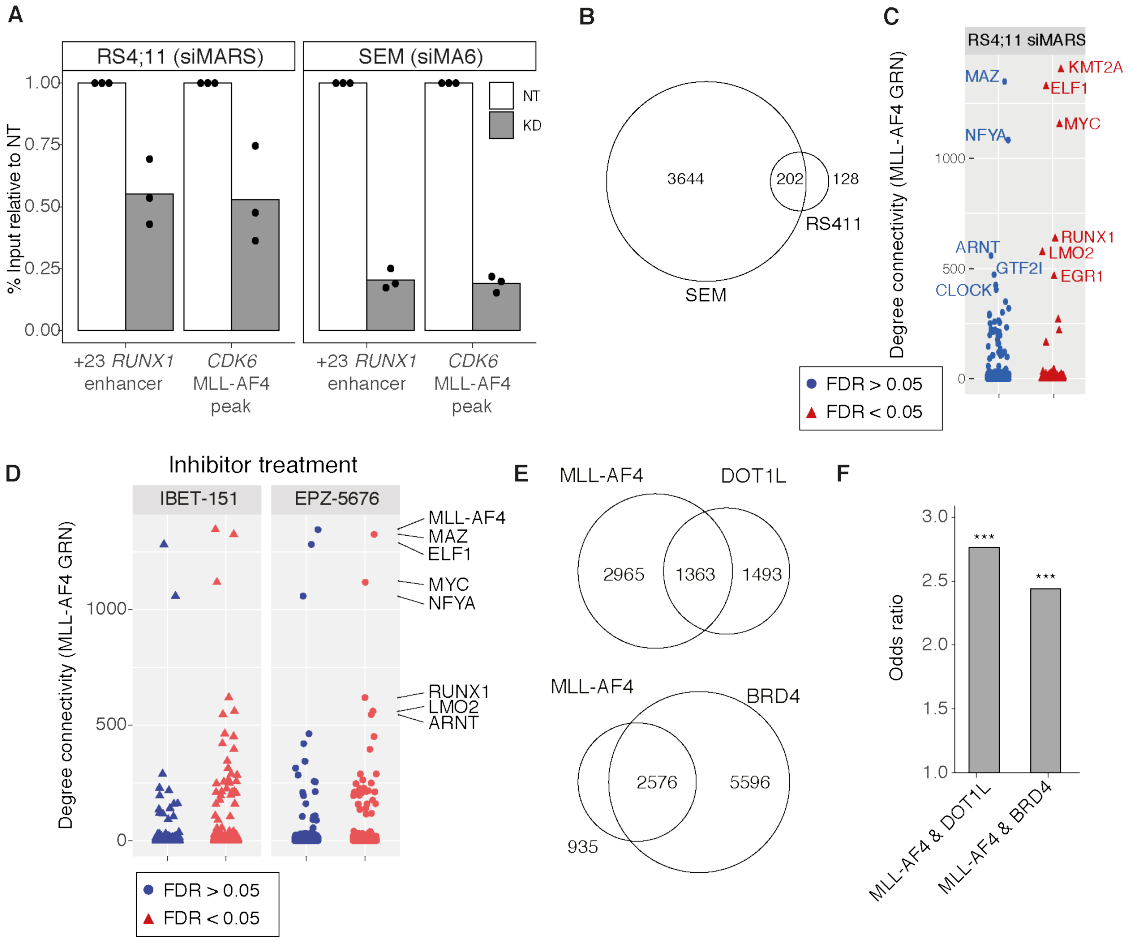


Figure 4.4: Validation of the MLL-AF4 GRN. (A) MLL-N ChIP-qPCR data in RS4;11 and SEM cells after *MLL-AF4* KD, normalized to input chromatin and displayed relative to non-targeting (NT) control. siMA6 and siMARS refer to SEM and RS4;11 specific *MLL-AF4* siRNA, respectively (Thomas *et al.*, 2005). **(B)** Overlap of nodes between SEM GRN and RS4;11 *MLL-AF4* KD DEGs. **(C-D)** Number of DEGs after *MLL-AF4* KD in RS4;11 cells ($n = 3$) (C), or 1.5 hours' IBET and 7 days' EPZ-5676 treatment ($n = 3$) (D), plotted against degree centrality in the SEM MLL-AF4 GRN. Blue points: FDR > 0.05, red points: FDR < 0.05. **(E)** Overlap of nodes between MLL-AF4 GRN, and IBET-151 and EPZ-5676 treatment DEGs. **(F)** Fisher's exact test of overlaps shown in (E), showing odds ratios. (***) P < 0.001. *Raw data for D-F was generated previously in the Milne lab, and analysed by me. Adapted from Harman* et al.*, 2021.*

this reduced network many of the most central GRN nodes were still affected upon RS4;11 *MLL-AF4* KD, including *RUNX1*, suggesting that these central factors are highly sensitive and consistent across MLL-AF4 ALL cell lines (Fig. 4.4C).

To further validate the GRN, I tested whether the GRN nodes reflect activity of MLL-AF4 complex components. DOT1L is a core complex factor (Bernt *et al.*,

2011; Biswas *et al.*, 2011; Krivtsov *et al.*, 2008; Lacoste *et al.*, 2002; Lin *et al.*, 2016; Milne *et al.*, 2005b; Mueller *et al.*, 2007; Okada *et al.*, 2005; Slany, 2020) that should strongly overlap with the GRN, and BRD4 is an associated cofactor with a general role in gene activation (Dawson *et al.*, 2011; Zuber *et al.*, 2011b) that is not specific to the MLL-AF4 complex. Comparing the MLL-AF4 GRN with DEGs following 7 days' EPZ-5676 (DOT1L inhibitor) treatment and 1.5 hours' IBET-151 (BRD4 inhibitor) treatment (Published data from Godfrey *et al.*, 2019 and Crump *et al.*, 2021), both inhibitors overlapped with the GRN and impacted central TFs (Fig. 4.4D-E). DOT1L inhibition impacted less of the MLL-AF4 GRN than BRD4, but with greater specificity (Fig. 4.4F). This analysis serves to validate that the MLL-AF4 GRN represents aspects of BRD4 and DOT1L activity and is therefore a reliable representation of MLL-AF4 complex behaviour.

4.3 Integrating patient expression data highlights regulatory modules, and a set of "core" MLL-FP TFs

MLL translocations can involve many different fusion partners, and can result in either ALL or AML. There are also cases where lineage switching occurs after relapse (Dorantes-Acosta *et al.*, 2012; Gardner *et al.*, 2016). This suggests similarity in the behaviour across different MLL-FP complexes, in both ALL and AML lineages, and suggests there may be a core set of targets driven by MLL-FPs more generally. Further, the translocation events often occur in utero resulting in infant and childhood leukaemias (Ford *et al.*, 1993; Greaves, 2005, 2018; Jackson *et al.*, 2021), and fetal programs can be overexpressed in *MLL*r leukaemia (Rice *et al.*, 2021), indicating that the MLL-AF4 GRN originates in a fetal regulatory landscape. To assess whether the GRN model represents common ALL and AML biology, or shares characteristics with fetal B-cell development, I overlapped the GRN with patient and fetal bone marrow (FBM) expression data to create sub-networks (Fig. 4.5A). Expression data was taken from *MLL*r ALL patients (Agraz-Doblas *et al.*,

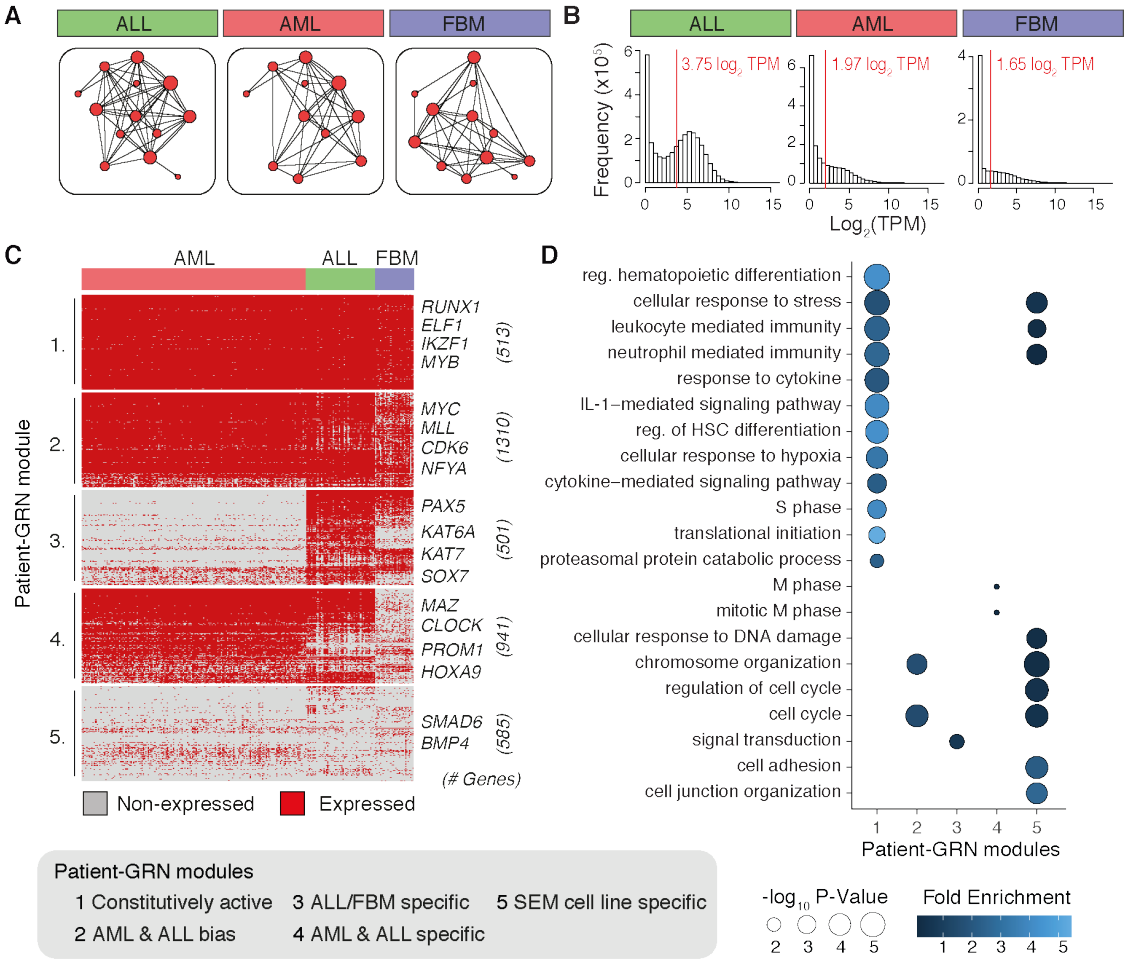


Figure 4.5: The MLL-AF4 GRN can be split into regulatory modules based on patient gene expression. **(A)** Schematic illustration of sub-networks reflecting expressed genes in patient samples. **(B)** Histograms of $\log_2$ TPM expression in AML, ALL and FBM datasets. Red line indicates threshold for defining active gene expression (mean $\log_2$ TPM). **(C)** GRN node activity across sub-networks as defined in B. Genes were grouped into patient-GRN modules using k-means. Rows represent genes; columns represent samples. **(D)** GO biological process enrichment for patient-GRN modules. *Normalised expression data for this analysis was sourced from published data, with GRN analyses performed by me. Adapted from Harman* et al.*, 2021.*

2019), AML patients with various chromosomal abnormalities (The Cancer Genome Atlas Research Network, 2013), and HSPCs and B cells from normal FBM (O'Byrne *et al.*, 2019). These FBM populations included HSCs, multipotent progenitors (MPP), lympho-myelo primed progenitors (LMPP), pre-pro-B and pro-B cells. A threshold was used to define active genes, based on the mean gene expression within each dataset. Active genes were used to filter the MLL-AF4 GRN into patient specific sub-networks (Fig. 4.5B, methods section 2.15.3, p.57). Median expression

was unusable due to low sequencing depth in the FBM data.

4.3.1 Regulatory modules of the GRN

A binary matrix of gene activity across sub-networks was generated, then clustered into five patient-GRN modules (Fig. 4.5C). Note that *MLL* expression in the case of *MLL*r leukaemias also includes MLL-FP activity. A caveat of this analysis is that levels of expression are not addressed, which may have more subtle but important consequences. For example, *RUNX1* is present in module 1 and active across all samples, however *RUNX1* has been shown to be overexpressed in MLL-AF4 ALLs compared to MLL-AF9 AML and is important for leukaemic growth (Wilkinson *et al.*, 2013). Thus, expression level rather than simply just presence of gene expression is likely an important indicator of functional importance. Nevertheless, this analysis serves to highlight modules that are potentially important across multiple leukaemias.

Nodes in modules 1 and 2 are active across all leukaemias and FBM populations and are associated with haematopoietic differentiation as a key process, emphasising a maintained fetal haematopoietic program (Fig. 4.5D). Module 3 is active in ALL and a subset of FBM samples, and identifies *SOX7* and *KAT6A* as ALL-specific, and *PAX5* and *KAT7* in both ALL and FBM (Fig. 4.5C). The lack of AML activity implies this module differentiates ALL and AML, reinforced by the presence of *PAX5*, a critical regulator of B lymphopoiesis (Medvedovic *et al.*, 2011), and *SOX7*, which is highly expressed in ALL over AML and drives proliferation (Cuvertino *et al.*, 2016). Module 4 is instead active in both AML and ALL datasets, with low activity in FBM samples (Fig. 4.5C), indicating nodes distinct from the normal B progenitor program. Interestingly, this includes the central GRN node *MAZ*, indicating it is a leukaemia specific target. *PROM1* (CD133) is also in this cluster, which is normally expressed in more stem and progenitor cells but can be dramatically upregulated by MLL-FPs and is essential for many ALL leukaemias (Godfrey *et al.*, 2021; Mak *et al.*, 2012; O'Byrne *et al.*, 2019). Module 5 is inactive

across all datasets (Fig. 4.5C), and pathway enrichment shows terms related to cell junction organisation and adhesion, response to stress, and DNA damage response (Fig. 4.5D), which likely represents an immortalised cell program. Together, this analysis breaks down the MLL-AF4 GRN into different modules representing ALL specific, core AML and ALL, and B-lymphopoiesis co-opted circuits. ALL specific nodes are expected as an ALL based GRN, while nodes active in both ALL and FBM populations are suggestive of a B progenitor program maintained in ALL, which may be overactivated by MLL-AF4. A particularly interesting subset are nodes active in both ALL and AML. This is indicative of a program active in both lineages, and as *MLL*r may undergo lineage switching, this may represent a consistent program that enables this switch to occur.

4.3.2 High centrality nodes represent a core MLL-FP program

As MLL-AF4 may regulate a core set of nodes across lineages (Fig. 4.5C), it is important to establish which TFs drive this core program. The most central nodes of the SEM MLL-AF4 GRN (degree centrality > 500) are active across all ALL and AML patient sub-networks (Fig. 4.6A), suggesting these TFs play a common role across lineages. Central GRN nodes are considerably less conserved in the fetal B progenitor program (Fig. 4.6A), suggesting that this is regulated by distinct TFs, although this could be a consequence of the low sequencing depth or diversity of fetal populations. Patient-GRN modules 1, 2 and 4 represent the core regulatory logic between acute leukaemias (Fig. 4.5C), and show strikingly higher centrality compared to modules 3 and 5 (Fig. 4.6B). Therefore, the central factors of these clusters can be taken as a panel of core TFs applicable to both AML and ALL (Fig. 4.6C).

The analysis so far has identified highly connected TFs present across AML and ALL lineages. This has focused on the MLL-AF4 centred GRN, but our analysis could be extended to other MLL-FPs, such as MLL-AF9, to determine if there

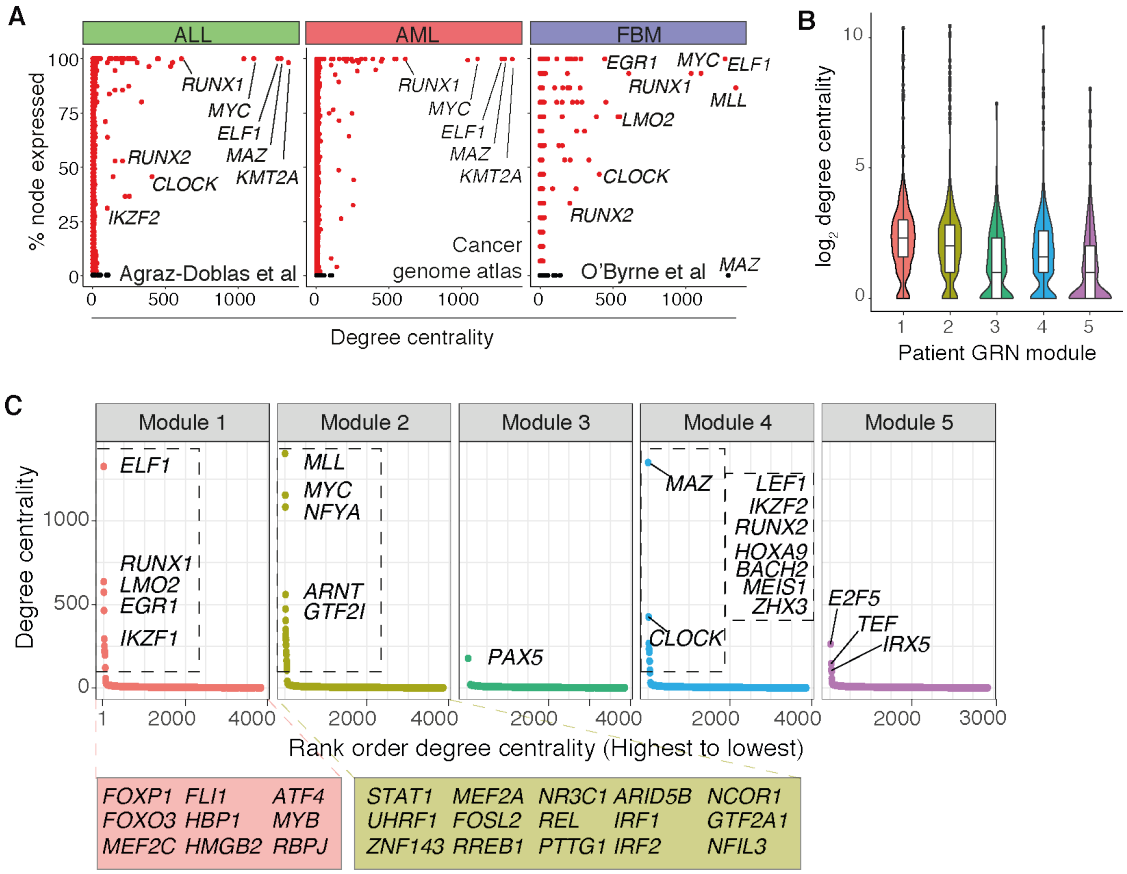


Figure 4.6: Patient expression modules reveal a set of core, central nodes. (A) Relationship between percentage of patient samples expressing each gene and MLL-AF4 GRN centrality. **(B)** Distribution of $\log_2$ degree centrality of GRN nodes, stratified by patient-GRN module. **(C)** Top-most GRN nodes by degree centrality, stratified by patient-GRN modules. Annotated nodes in Clusters 1, 2, and 4 represent the core TFs of the GRN model. *Adapted from Harman* et al., *2021.*

is a general MLL-FP core network. To test this, I compared MLL-N ChIP-seq profiles between SEM and THP-1 cells (MLL-AF9 AML). Of SEM MLL-AF4 bound genes in the MLL-AF4 GRN, 66% are bound by MLL-AF9 in THP-1 cells (Fig. 4.7A). To determine whether this commonality extends to central targets of MLL-AF4, I performed the same comparison with RUNX1 ChIP-seq. Interestingly, only 35% of SEM RUNX1 bound genes are in common with THP-1, suggesting that while there is common MLL-FP behaviour the central TFs are lineage specific (Fig. 4.7A). MLL-AF4 and MLL-AF9 binding profiles are broadly similar at common gene targets (Fig. 4.7B), with similar enhancer occupancy and MLL-FP spreading (Kerry *et al.*, 2017). The core TF program is largely bound by MLL-FP in both

SEM and THP-1 cells, with some exceptions including *MAZ*, *NFYA* and *EGR1* (Fig. 4.7D). In contrast, RUNX1 binding profiles at enhancers differ, such as at the *GFI1* locus (Fig. 4.7C). Despite these differences, many of the central GRN factors are bound by RUNX1 in both cell types (Fig. 4.7E). These analyses validate the capacity of the core MLL-FP TFs to regulate both ALL and AML leukaemias, using different MLL fusion partners, and highlight differences in RUNX1 binding between lineages. The core program of these leukaemias are similar, but the wider regulatory network may be context dependent.

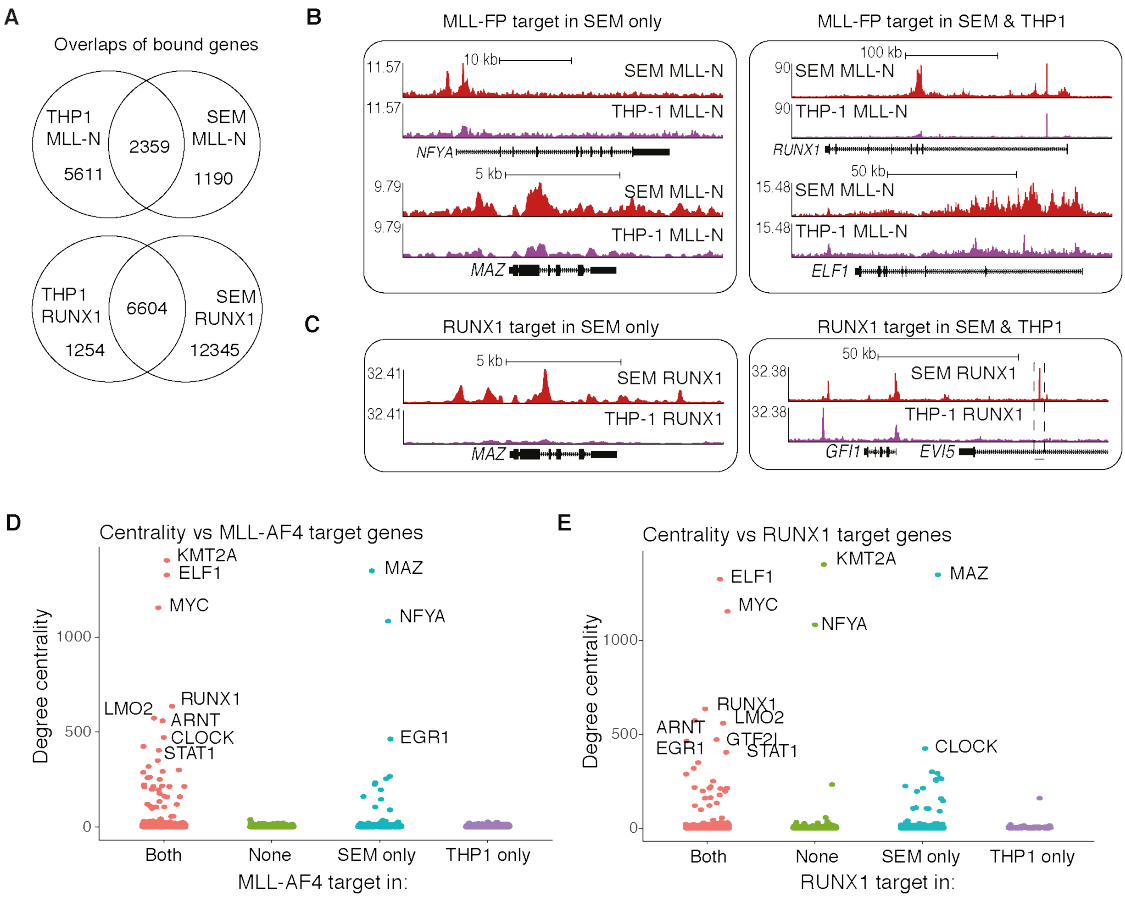


Figure 4.7: The MLL-AF4 GRN predicts core interactions representative of both MLL-AF4 ALL and MLL-AF9 AML. (A) Overlaps of MLL-FP or RUNX1 bound genes in THP-1 and SEM cells. **(B, C)** ChIP-seq tracks for MLL-N (B) and RUNX1 (C) in SEM and THP-1 cells. Dashed box highlights SEM specific RUNX1 peak. **(D, E)** Association of degree centrality with MLL-FP (D) or RUNX1 (E) bound gene overlaps in THP-1 and SEM cells as shown in (A). *Adapted from Harman* et al., *2021.*

4.4 Validating the importance of central regulatory hubs

The analysis thus far builds a case for a core program common across *MLL*r leukaemia, and implicates several TFs as key regulatory hubs based on degree centrality. Previous studies in *S. cerevisiae*, *C. elegans*, and *D. melanogaster* have shown that multiple centrality measures are associated with essential proteins (Hahn *et al.*, 2005; Jeong *et al.*, 2001; Koschützki *et al.*, 2008). However, while important, many highly connected TFs in networks have overlapping roles, promoting robustness and suggesting multiple perturbations may be required to disrupt a process (Byrne *et al.*, 2007; Lehner *et al.*, 2006; MacNeil *et al.*, 2011). As such, it is important to confirm that centrality measures can be used to determine important nodes, and estimate the nodes most susceptible to perturbation.

4.4.1 Essentiality screens validate the importance of several central TFs

To validate whether central TFs are important for leukaemogenesis, I integrated publicly available CRISPR essentiality screens. The Cancer Dependency Map (DepMap, Avana 21Q1 dataset, Doench *et al.*, 2016; Meyers *et al.*, 2017) found *MYC* to be pan-essential, *RUNX1* to be essential in multiple haematopoietic cancer models, and *ELF1* and *MAZ* to be non-essential (Fig. 4.8A). As this is a large and generalised screen, I integrated a more targeted screen from Tzelepis *et al.*, 2016 which includes two *MLL*r cell lines, MOLM-13 (MLL-AF9 AML) and MV4-11 (MLL-AF4 AML), alongside two non-leukaemic cancer cell lines, HT-29 (colon adenocarcinoma) and HT-1080 (fibrosarcoma). To find genes essential specifically in leukaemia, hits were categorised as non-essential, non-specific essential (hit in HT-29 or HT-1080), or essential in one or both *MLL*r leukaemias (hit in MOLM-13 and/or MV4-11, but not HT-29 or HT-1080) (Fig. 4.8B-C). In line with the DepMap data, *MYC* was pan-essential, while *MAZ* and *ELF1* were not essential in any cell line. This suggests *MAZ* and *ELF1*, while highly connected, are not

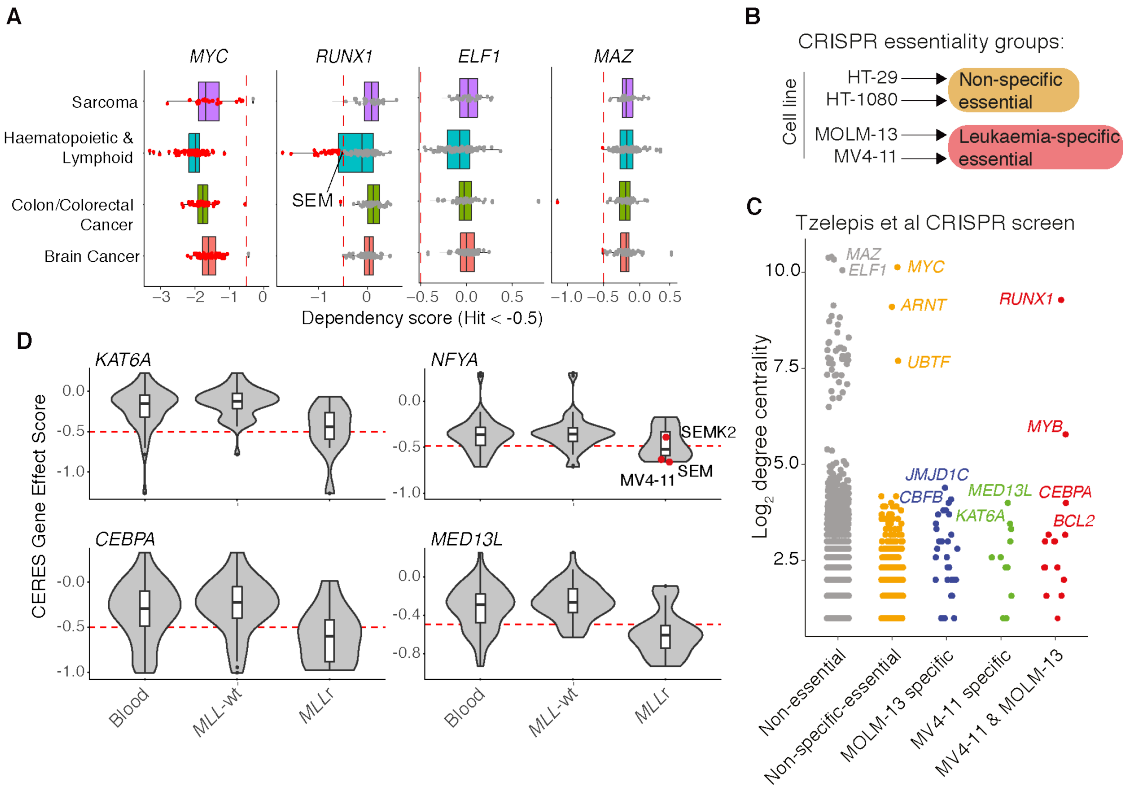


Figure 4.8: CRISPR essentiality screens highlight RUNX1 as a critical central TF in MLL-FP leukaemias. (A) CRISPR essentiality CERES scores from DepMap (Avana 21Q1, Doench *et al.*, 2016; Meyers *et al.*, 2017). Genes are considered essential with a dependency score below -0.5. Each datapoint represents a different cell line, with essential hits highlighted red. **(B)** Illustration for how CRISPR essentiality screen hits from Tzelepis *et al.*, 2016 are categorized. Essential genes in HT-29 or HT-1080 were "non-specific essential". Genes not essential in HT-29 or HT-1080 but essential in MOLM-13 or MV4-11 were specific to leukemia. **(C)** Stratification of $\log_2$ degree centrality (MLL-AF4 GRN) by CRISPR essentiality groups as outlined in B. **(D)** CERES scores in haematopoietic cancer cell lines from DepMap, stratified by MLL translocation status. Red datapoints for *NFYA* indicate MLL-AF4 cell lines. *Analysis in D performed by T. Wilson, supervised by me. CRISPR scores sourced from DepMap and Tzelepis* et al., *2016. Adapted from Harman* et al., *2021.*

important for survival or proliferation, at least in AML. *RUNX1* was essential in both MOLM-13 and MV4-11 cells, highlighting its importance under the control of both MLL-AF4 and -AF9 fusion proteins. Other central leukaemia-specific hits include *MYB*, *MED13L*, *CEBPA*, and *KAT6A* (Fig. 4.8C).

The MV4-11 and MOLM-13 essential gene *CEBPA*, along with the MV4-11 specific hits *MED13L* and *KAT6A*, show greater essentiality for *MLL*r samples over non-*MLL*r blood cancers (Fig. 4.8D). NF-YA is a central (but not core to all MLL-

FPs) node in the GRN and is not significantly essential in the Tzelepis *et al.*, 2016 screen. However, the DepMap screen shows higher essentiality for NF-YA in *MLL*r samples, with *MLL-AF4* cell lines (SEM and MV4-11) showing the high essentiality (Fig. 4.8D). This indicates that not only are these genes essential for MLLr AML, but show generally greater essentiality for *MLL*r leukaemias than other haematopoietic malignancies, possibly in part as a core-program targeted by all MLL-FPs. Overall, this analysis shows that degree centrality highlights many essential genes, and identifying high centrality in TFs critical for both MLL-AF4 and MLL-AF9 AML.

4.4.2 Stress is a better indicator of essentiality than degree centrality

As demonstrated, while degree centrality highlights nodes that regulate a wide range of targets and are often critical, this measure does not always translate into essentiality for leukaemic cells. Stress centrality (a variant of betweenness centrality) may better reflect essentiality, as it represents nodes less likely to be compensated for upon perturbation (see introduction section 1.2.5, p.11, Fig. 1.1). In fact, while there is a slight enrichment in degree centrality for MV4-11 essential genes, there is much higher stress centrality observed in MV4-11 essential genes over non-essential genes (Fig. 4.9A). *MAZ* was a non-essential TF, which despite high connectivity, shows relatively low stress centrality than comparably connected TFs (Fig. 4.9B). Exploring the ratio between stress and degree centrality, RUNX1 has a far higher stress for its connectivity than other TFs (degree 636, stress 80864), and while MYB is not highly connected it also shows a very high stress:degree ratio (degree 56, stress 3797) (Fig. 4.9C). It has been suggested that nodes with low degree and high stress are particularly important factors for connecting network modules (Joy *et al.*, 2005; Koschützki *et al.*, 2008), placing RUNX1 as a critical TF in bridging interactions between factors, such as MLL-AF4.

As discussed, central TFs may contain redundant regulatory logic that compensates for specific perturbation. To investigate alternative nodes that may facilitate

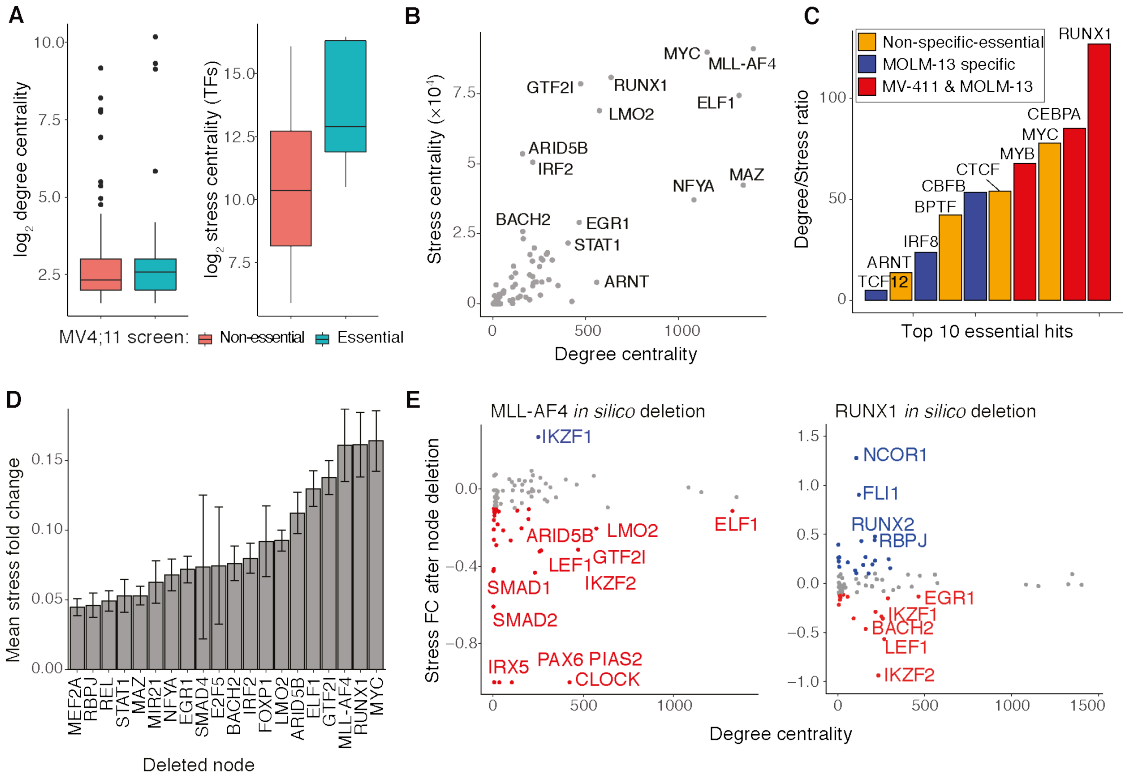


Figure 4.9: Centrality analysis of MLL-AF4 GRN. (A) Log_2 degree and stress centrality of the MLL-AF4 GRN nodes plotted by MV4-11 essentiality from Tzelepis *et al.*, 2016. **(B)** Relationship between degree and stress centrality of MLL-AF4 GRN nodes. **(C)** Rank order plot of ratios between stress and degree centrality for ten most essential GRN hubs from Tzelepis *et al.*, 2016. **(D)** Top 20 in silico deleted nodes by mean stress FC (methods section 2.15.5, p.59). **(E)** Stress FC after in silico deletion of RUNX1 or MLL-AF4, plotted against degree centrality. Blue and red indicate positive (FC > 0.1) and negative (FC < -0.1) stress FC, respectively. *Adapted from Harman* et al., *2021.*

TF compensation, I established an approach where each GRN node was deleted in silico, and stress centrality values recalculated. Stress values before and after in silico deletion were compared as stress fold change (stress FC), and summarised for every node (Fig. 4.9D). RUNX1 and MYC in silico deletion showed the greatest stress redistribution, above MLL-AF4, while the non-essential factors MAZ and NFYA show less stress FC redistribution (Fig. 4.9D). Looking in detail, MLL-AF4 deletion results in negative stress FC while RUNX1 shows positive and negative stress FC. RUNX1 in silico deletion shows increased NCOR1, FLI1, RBPJ and RUNX2 stress, while factors such as IKZF1 and IKZF2 are reduced (Fig. 4.9E), indicating that these TFs provide an alternative route to reach RUNX1 target nodes.

RUNX2 is expected as it shares the RUNT domain with RUNX1, and therefore can likely compensate for perturbation. RUNX1 has been shown to cooperate with FLI1, altering its binding affinity for the FLI1 motif and remodelling FLI1 regulatory logic (Gilmour *et al.*, 2018; Lichtinger *et al.*, 2012). RBPJ is a transcriptional repressor, yet forms a transactivating complex with NOTCH intracellular domain (NICD) (Kojika *et al.*, 2001). RUNX1 and RBPJ-NICD cooperate at specific loci, for example to coregulate the +1.27 Mb *MYC* enhancer and *IL7R* enhancers (Choi *et al.*, 2017; Wang *et al.*, 2014). This cooperative behaviour is indicative of shared regulatory logic, and while these factors function in combination, it is likely they regulate common sites in the absence of RUNX1. Genes that decrease in stress may reflect a breakdown in network pathways. For example, if node pairs must communicate through a RUNX1:IKZF1 axis then loss of RUNX1 will break down the path of communication via IKZF1. This places IKZF1 as a key target of RUNX1, required for downstream communication within the network. This lines up with Runx1 regulation of Ikzf1 in EHT, which was predicted to enact TF cascades to repress Notch genes (Fig. 3.14D, p.91). This analysis serves to rank nodes based on stress centrality, which may better reflect importance, and predicts compensatory mechanisms upon perturbation. In particular it highlights RUNX1, MYB, CEBPA and MYC as highly important TFs, and notes potential redundancy between RUNX1 and specific factors at a subset of targets.

4.5 MLL-AF4 and RUNX1 cooperate to regulate gene expression

4.5.1 Generating, and characterising a RUNX1 driven GRN model of leukaemia

To begin to address how MLL-AF4 regulates a wider network in combination with TFs, I selected RUNX1 as an example of a critical MLL-AF4 activated node.

Across multiple analyses RUNX1 was promoted as a critical TF within the MLL-AF4 GRN, it is highly central and is a core-TF active in both MLL-AF4 ALL and MLL-AF9 AML. To understand RUNX1 regulatory logic in detail, I established a RUNX1-centred GRN. Nascent RNA-seq data following 96 hours' *RUNX1* KD

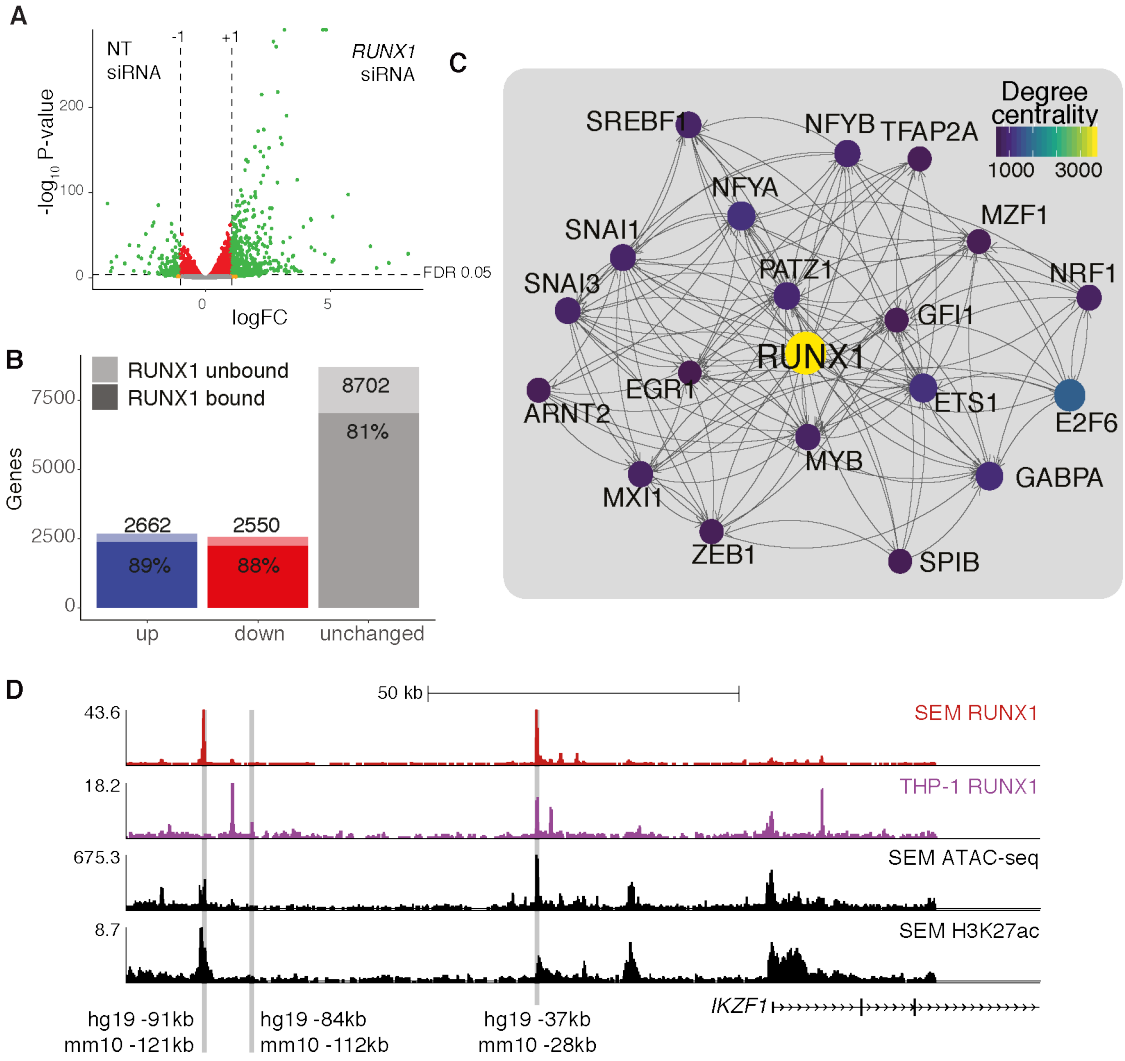


Figure 4.10: Establishing a RUNX1 centred GRN in SEM cells. (A) Relationship between -$\log_{10}$ P value and logFC in *RUNX1* KD nascent RNA-seq. Genes are coloured by logFC and FDR thresholds as indicated by dashed lines. **(B)** DEGs from nascent RNA-seq after 96 h *RUNX1* KD ($n = 3$). Shaded bar represents RUNX1-bound genes. **(C)** The top 20 genes of the RUNX1 GRN by degree centrality. Lines indicate predicted interaction from protein to gene locus, with arrowheads pointing downstream. **(D)** ChIP-seq and ATAC-seq tracks at the human *IKZF1* locus. Mouse -28, -112, and -121 enhancers are highlighted, matched by sequence similarity using UCSC BLAT. *Raw nascent RNA-seq data was generated by M. Tapia (Milne lab) and analysed by me. Adapted from Harman* et al., *2021.*

generated by M. Tapia revealed 5212 DEGs, the majority of which were bound by RUNX1 (Fig. 4.10A-B). A RUNX1-centric GRN was established using this nascent RNA-seq and previously published RUNX1 ChIP-seq (Wilkinson *et al.*, 2013), using the same approach as in Fig. 4.3. Many known targets of RUNX1, such as *GFI1* (Wilson *et al.*, 2010) and *MYB* (Choi *et al.*, 2017), were central to the RUNX1 GRN (Fig. 4.10C). A number of Runx1 targets in the EHT GRN (see section 3.4.3, p.81) are also present here, including, among others, *MYB*, *IKZF1/2*, *SNAI1/3*, *ZEB1/2*, *GFI1* and the NOTCH targets *HES1*, *HEY1* and *HEY2*. I compared RUNX1 binding at the *IKZF1* locus with the EHT GRN and the mouse predicted enhancer elements (Fig. 3.15, p.93). RUNX1 binding at *IKZF1* shows similar occupancy of the -28 and -121 enhancer elements (-37 kb and -91 kb in human) as in EHT, though the -112 enhancer (-84 kb in human) is not bound by RUNX1 (Fig. 4.10D). This suggests both GRN models capture common RUNX1 regulatory logic, though with some differences in enhancer occupancy, that is preserved across species and between different cell contexts.

4.5.2 RUNX1 and MLL-AF4 GRNs intersect at leukaemogenic processes

The central hypothesis of this chapter is that MLL-AF4 regulates a wider network through driving RUNX1 expression (Fig. 4.1B), and the analysis so far has identified the TF RUNX1 as a key downstream effector of MLL-AF4 activity. Genes that are sensitive to both *MLL-AF4* and *RUNX1* KD in the GRN models (2279 DEGs, Fig. 4.11A) were enriched for processes related to haematopoiesis, regulation of cell death, and B cell proliferation, consistent with an ALL expression program (Fig. 4.11C). Pathway enrichment of these co-regulated genes revealed terms for epigenetic regulation such as DNA and histone methylation, cell cycle, and RUNX1 regulation (Fig. 4.11C). These analyses suggest that the intersection of MLL-AF4 and RUNX1 GRNs drive a pro-leukaemic program.

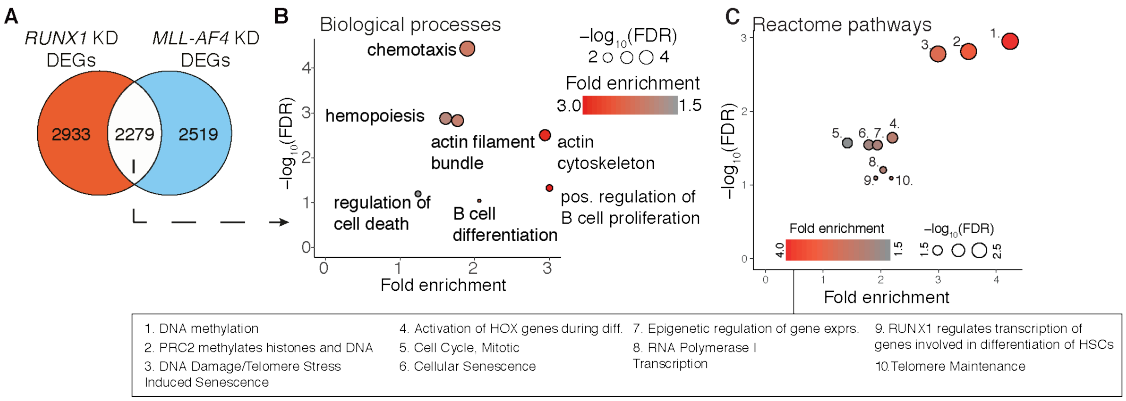


Figure 4.11: The overlap between MLL-AF4 and RUNX1 GRNs is enriched for processes promoting leukaemogenesis. (A) Overlap between *MLL-AF4* KD DEGs (Fig. 4.1A) and RUNX1 KD DEGs (Fig. 4.10A-B). **(B-C)** GO biological process (B) and reactome pathway (C) enrichment for overlap shown in A. *Adapted from Harman* et al.*, 2021.*

4.6 MLL-AF4 and RUNX1 direct both FFL and TF cascade regulation

4.6.1 Identifying FFL and cascade regulatory circuits

While MLL-AF4 and RUNX1 regulatory logic intersects at key pathways, it is yet unclear whether these factors cooperate directly, or if this intersection is driven through RUNX1 acting as an intermediate TF. To dissect how RUNX1 and MLL-AF4 co-regulate gene targets I integrated the GRN models, then identified FFL and TF cascade regulatory motifs. Regulatory logic was predicted using the siRNA perturbation data (Fig. 4.1A, Fig. 4.10A-B), where loss of gene expression upon KD implies positive regulation, and vice versa. This method assumes expression responses are a direct result of MLL-AF4 or RUNX1 regulation. It is well established that MLL-AF4 drives *RUNX1* overexpression (Wilkinson *et al.*, 2013), so this constrains possible FFL subtypes into just four categories. These include two coherent (C1- and C3-FFL) and two incoherent FFLs (I1- and I3-FFL), with subtypes classified by MLL-AF4 and RUNX1 regulation of genes (Mangan *et al.*, 2003a) (Fig. 4.12A). Cascade motifs were identified as genes directly regulated by RUNX1, but indirectly (i.e., not bound) by MLL-AF4. These were further classified into so-called "coherent" and "incoherent" groups (Fig. 4.12A, coherent

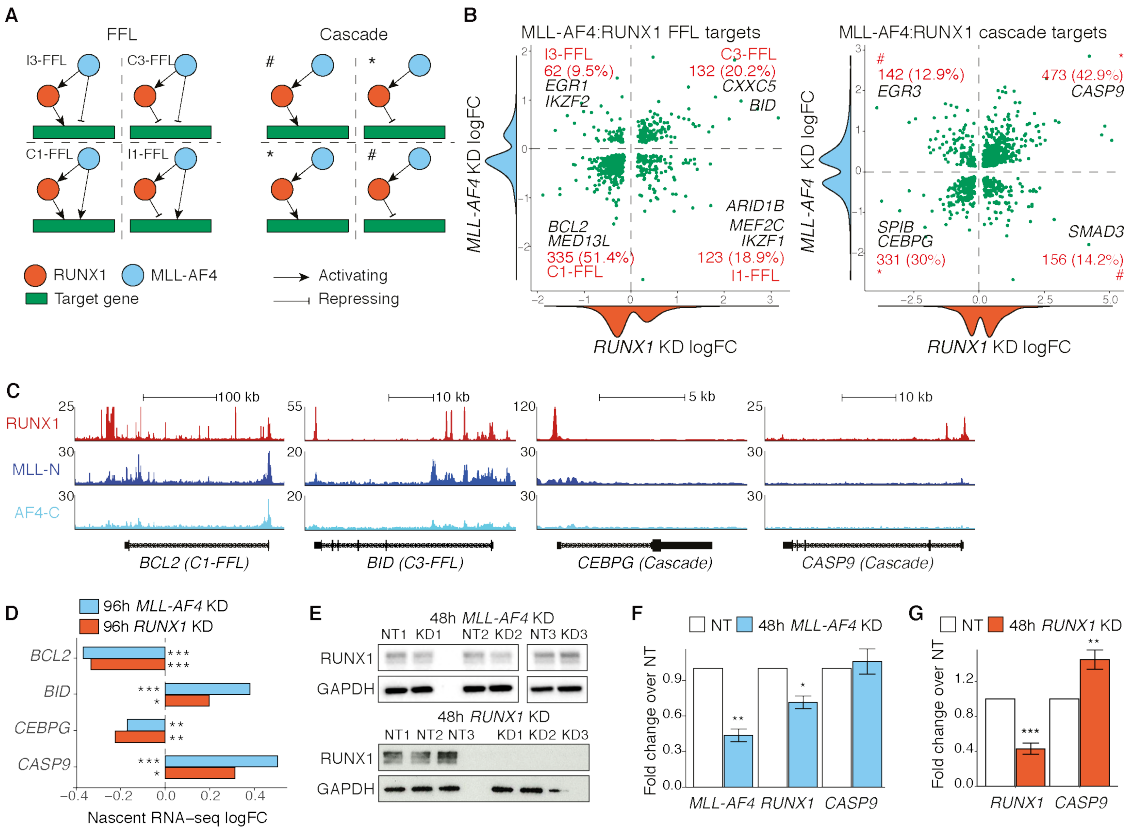


Figure 4.12: Network motifs describing interplay between RUNX1 and MLL-AF4. (A) Illustration of FFL sub-types (left) and cascade (right) motifs. *: Same sign of effect. #: Opposing sign of effect. **(B)** LogFC response to *RUNX1* and *MLL-AF4* KD in FFL (left) and TF cascade (right) target genes. Density plots of logFC distribution along axis. Quadrants of scatter plots align with FFL and cascade subtypes shown in A. **(C)** ChIP-seq tracks for MLL-N, AF4-C, and RUNX1 at *BCL2* (C1-FFL target), *BID* (C3-FFL target), *CEBPG* and *CASP9* (cascade targets). **(D)** Nascent RNA-seq logFC following 96h *RUNX1* or *MLL-AF4* KD. **(E)** Western blot for RUNX1 protein in SEM cells after 48 h *MLL-AF4* KD or 48 h *RUNX1* KD, with GAPDH as loading control. **(F, G)** qRT-PCR assaying *MLL-AF4*, *RUNX1* and *CASP9* expression following 48 h *MLL-AF4* KD (F) or 48 h *RUNX1* KD (G) in SEM cells ($n = 3$). (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.001$ *Adapted from Harman* et al.*, 2021.*

and incoherent denoted by * and # respectively). Coherent cascades show an alignment between the effect of *RUNX1* and *MLL-AF4* KD, which agrees with the hypothesis that RUNX1 targets a wider network as an intermediate for MLL-AF4.

Overall, we found that 71.6% of FFLs were coherent, suggesting that MLL-AF4 and RUNX1 act cooperatively at the same loci (Fig. 4.12B). For example, the key MLL-AF4 target *BCL2* is activated in a C1-FFL, while *BID* is repressed in

a C3-FFL, with both loci bound by MLL-AF4 and RUNX1 (Fig. 4.12C-D). FFL subtypes are typically biased towards a C1-FFL configuration (Joanito *et al.*, 2018; Mangan *et al.*, 2006), which is recapitulated in our analysis, with 51.5% of FFLs showing cooperative activation of target genes (Fig. 4.12B), further suggesting that RUNX1 at these FFL sites is biased towards activation. Interestingly, the EHT Runx1 target *Ikzf1* (Fig. 3.10, p.82) is targeted in an I1-FFL in the MLL-AF4 GRN, where RUNX1 appears to repress *IKZF1* rather than activate (RUNX1 KD logFC: 0.93, FDR: 2.8×10^{-43}), despite similar RUNX1 binding profiles to EHT (Fig. 4.10D). This change in RUNX1 regulation could be due to the leukaemic or the B cell context, in the presence of different cofactors and TFs.

The majority of cascade motifs are "coherent" (72.9%, Fig. 4.12B), suggesting that RUNX1 is a strong determining factor for regulation at the majority of cascade motifs. Example cascade targets are *CEBPG* (activated) and *CASP9* (repressed), where RUNX1 is bound but MLL-AF4 is not (Fig. 4.12C-D). Incoherent cascades (27.1%, Fig. 4.12B) are likely a result of noise in the nascent RNA-seq experiments or regulatory control by additional TFs. Interestingly, while FFLs show a strong activation bias, cascade motifs are balanced between activating and repressive. This suggests that while RUNX1 behaves as an activator when at the same genes as MLL-AF4, at other sites RUNX1 may act as either an activator or a repressor. This is consistent with dual RUNX1 activity in haematopoietic differentiation (Canon *et al.*, 2003; Kuvardina *et al.*, 2015; Lutterbach *et al.*, 2000).

The repressed cascade target *CASP9* (caspase-9) was validated by performing siRNA KD for a shorter length of time. To generate the GRN, *MLL-AF4* and *RUNX1* KDs were performed for 96 hours to achieve a robust reduction in protein, resulting in significant *CASP9* upregulation (Fig. 4.12D); however, this timepoint also favours secondary effects. A shorter-term *MLL-AF4* KD for 48 hours reduces *RUNX1* mRNA levels, but does not impact RUNX1 protein, and has little effect on *CASP9* expression ($P = 0.64$, Fig. 4.12E-F), consistent with indirect regulation of *CASP9* by MLL-AF4. *CASP9* is upregulated with just 48 hours' *RUNX1*

KD, as with 96 hours' KD, suggesting that RUNX1 may directly repress the gene (Fig. 4.12E, G). As *CASP9* is sensitive to *RUNX1* perturbation at an earlier timepoint, but not with *MLL-AF4*, this confirms that MLL-AF4 may act in this TF cascade through RUNX1 activity.

In publishing this data, in collaboration with M. Jamilly (Fulga lab), the MLL-AF4:RUNX1:*CASP9* TF cascade was implicated as a potential promoter of venetoclax resistance (Harman *et al.*, 2021). Venetoclax is a BCL2 inhibitor that is used to treat *MLL*r leukaemia (Khaw *et al.*, 2016; Vandenberg *et al.*, 2013), though drug resistance is a concern. As caspase-9 lies downstream of BCL2 in the apoptotic pathway, the MLL-AF4:RUNX1:*CASP9* cascade may provide a potential circuit for deterring apoptosis. To test whether RUNX1 activity interacts with venetoclax efficacy, I treated SEM cells with a small molecule RUNX1 inhibitor (active AI-10-104, inactive AI-4-88) that disrupts RUNX1 binding to the TF CBFβ (Illendula *et al.*, 2016; Illendula *et al.*, 2015), resulting in reduced RUNX1 DNA binding and expression (Fig. 4.13A-B). This reduced *RUNX1* expression is likely due to loss of positive RUNX1 autoregulation (Martinez *et al.*, 2016). I generated venetoclax dose-response curves, and co-treated with 10 μM RUNX1 inhibitor. Co-treatment with AI-10-104 yielded a lower venetoclax IC50 (78 nM) over AI-4-88 or DMSO treatment (208.97 nM, 193.47 nM) (Fig. 4.13C). This preliminary analysis suggests an interaction between venetoclax and RUNX1 activity, though *CASP9* levels appear unaffected by RUNX1 inhibition, so it is possible that this effect is mediated through alternative pathways (Fig. 4.13B). Additional replicates are needed to determine *CASP9* sensitivity to AI-10-104 treatment. Recent studies have noted improved venetoclax response in patients with *RUNX1* mutations, and have suggested this effect is through generally reducing the apoptotic threshold (Cherry *et al.*, 2021; Chow *et al.*, 2021). A more robust experimental design is required to determine if the effect is synergistic.

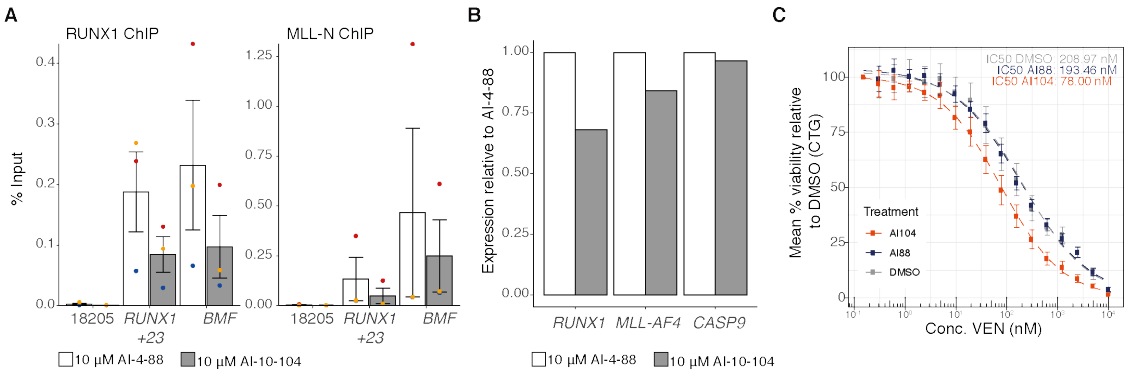


Figure 4.13: RUNX1 inhibitor treatment reduces the IC50 concentration of venetoclax. (A) RUNX1 and MLL-N ChIP-qPCR after 24 h 10 μM AI-10-104 or AI-4-88 (inactive control) treatment (as used in SEM cells in Illendula *et al.*, 2016), normalised to input chromatin. Regions probed are the *RUNX1* +23 enhancer and a RUNX1 bound peak at the *BMF* locus. 18205 refers to a negative control locus. Bars show mean, and error bars represent SEM ($n = 3$). **(B)** qRT-PCR analysis showing *RUNX1*, *MLL-AF4*, and *CASP9* expression after 24 h 10 μM AI-10-104 or AI-4-88 (inactive control) treatment ($n = 1$). Expression normalised to *GAPDH* shown relative to AI-4-88. **(C)** Dose response curve of SEM cells treated with venetoclax for 24 h. Cells were co-treated with 10 μM AI-10-104, AI-4-88 or DMSO). Viability measured using CTG, normalised to DMSO control. Dashed line represents dose response curve fit, and IC50 was calculated in the context of RUNX1 inhibitors. ($n = 8$).

4.6.2 Correlation between MLL-AF4 and RUNX1 ChIP-seq binding profiles

The number of MLL-AF4:RUNX1 FFLs and the shared regulatory logic at these genes suggests a synergistic interaction between the two factors. To investigate this possibility, I performed correlative analyses of genome wide chromatin binding profiles. For the TFs analysed (RUNX1 and MAZ), there was reasonable correlation with MLL-N ChIP-seq signal at promoters (Fig. 4.14A), and genome wide binding patterns show general co-localisation of the TFs and MLL-N (Fig. 4.14B). MLL-AF4 binding is broader than TFs peaks, and MLL-FP spreading is associated with highly overexpressed targets (Kerry *et al.*, 2017) (Fig. 4.14A), making precise comparisons between MLL-AF4 and TFs difficult. However, with increased MLL-AF4 spreading I observed a similar spread of TF binding, with an increased number of TF peaks. This is exemplified at *GNAQ*, where rather than typical spreading, a higher frequency of TF binding can be seen co-localising with MLL-AF4 spreading (Fig. 4.14A).

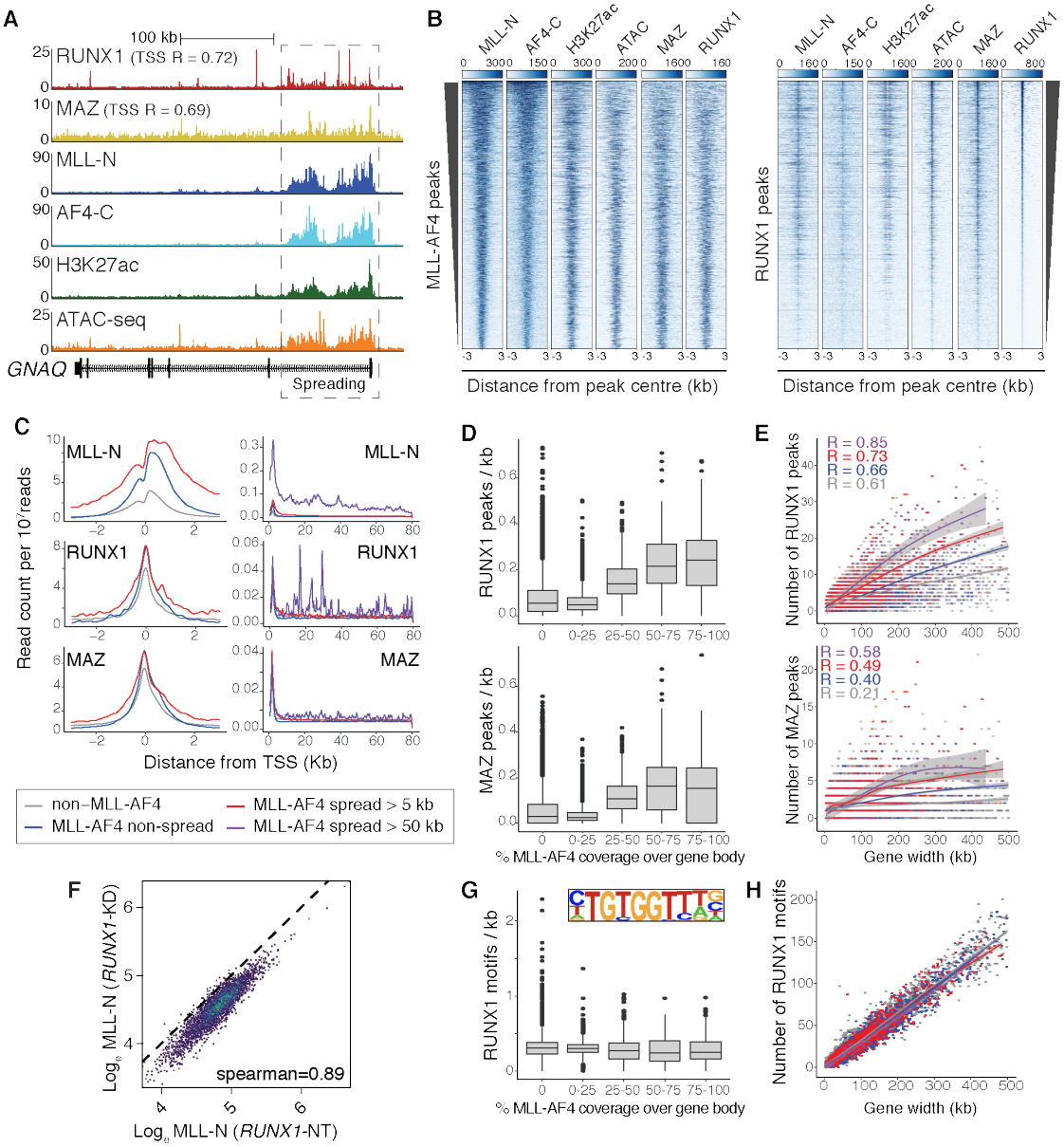


Figure 4.14: MLL-AF4 establishes an environment encouraging TF binding and FFL formation. (A) ChIP-seq tracks for GRN TFs, H3K27ac and ATAC-seq at *GNAQ*. Correlation (R) comparing TF ChIP-seq signal with MLL-N over expressed gene promoters. Dashed box highlights MLL-AF4 spreading. **(B)** ChIP-seq and ATAC-seq heatmaps in SEM cells, at MLL-AF4 peaks (left) and RUNX1 peaks (right), ordered by ChIP-seq signal. **(C)** Mean ChIP-seq signal at expressed promoters over a 6 kb (left) or 80 kb (right) window. Profiles stratified by MLL-AF4 binding classes as labelled. **(D,G)** TF ChIP-seq peaks/kb (D) or RUNX1 motifs/kb (G) over gene bodies, stratified by proportion of MLL-AF4 coverage. **(E,H)** Relationship between TF peak (E) or RUNX1 motif (H) frequency over gene bodies, and gene body length (kb), stratified by MLL-AF4 binding classes as in C. Local regression (LOESS) lines fit with 95% CI in grey. Correlation (R) calculated by MLL-AF4 binding classes. RUNX1 motif logo shown. **(F)** Log$_e$ MLL-N ChIP-seq reads at MLL-AF4 peaks following 48 hours' treatment with NT or RUNX1 siRNA.

To test whether increased TF binding is associated with MLL-AF4 genome wide, I analysed TF peak densities at gene loci. I first classified expressed genes by MLL-AF4 spreading status, differentiating between > 5 kb spreading (754 genes) and > 50 kb spreading (25 genes). TF binding at promoters was more intense at all MLL-AF4 targets over non-MLL-AF4, and showed increased signal outside the promoter at spreading targets. Further, with > 50 kb spreading there was increased TF signal over 50 kb away from gene promoters (Fig. 4.14C). The density of TF ChIP-seq peaks across a gene body (peaks/kb) increased with a greater proportion of MLL-AF4 coverage (Fig. 4.14D). The relationship between total TF peaks and gene body length is also stronger at MLL-AF4 spreading targets over non-spreading or unbound genes (Fig. 4.14E). The TF peak density at spreading targets is much higher with RUNX1 than MAZ, though the relationship is present in both cases. Together this suggests that with a greater spread of MLL-AF4 there is an increased frequency of TF binding.

As this analysis is correlative, whether TF binding encourages MLL-AF4 spreading or MLL-AF4 drives TF binding remains unclear. To determine whether TFs may influence MLL-AF4 binding or spreading patterns I first performed *RUNX1* siRNA KD. With 48 hours' *RUNX1* KD MLL-AF4 binding genome wide is slightly reduced, though profiles are largely maintained with no site-specific loss of MLL-AF4 and high correlation between KD and control samples (Fig. 4.14F). The slight reduction in MLL-AF4 binding could reflect a subtle impact on MLL-AF4 protein levels, though nascent *MLL-AF4* expression is unaffected (96h *RUNX1* KD RNA-seq: logFC = 0.1, FDR = 0.12). It remains difficult to disentangle the effect of *MLL-AF4* perturbation on RUNX1 binding, as MLL-AF4 directly drives *RUNX1* expression. Therefore, I cannot confidently conclude that MLL-AF4 directs TF binding, though the correlation is suggestive. Importantly, while RUNX1 binding frequency increases with MLL-AF4 spreading, RUNX1 motifs do not (Fig. 4.14G-H). This could imply the increased TF binding is through non-sequence specific DNA binding mechanisms, such as association with larger protein complexes, or that RUNX1 motifs are in

equal excess across genes and MLL-AF4 establishes increased RUNX1 motif binding affinity. This analysis suggests that MLL-AF4 and TFs co-localise to form FFL circuits, in a process that may be driven by MLL-AF4. An attractive hypothesis is that MLL-AF4 increases the binding affinity of TFs for existing motifs, through epigenetic modification and increased chromatin accessibility.

4.7 TF cooperation at MLL-AF4 targets drives expression and promotes leukaemogenesis

The data presented thus far offer a breakdown of MLL-AF4 and TF cooperation in FFLs, and indirect MLL-AF4 regulation using TFs as intermediates in cascade motifs. FFL motifs are particularly prominent, and there appears to be a relationship between MLL-AF4 and TF binding, promoting co-localisation at MLL-AF4 target sites. As MLL-AF4 and RUNX1 networks intersect at regulation of cell death and cell cycle processes (Fig. 4.11B), it is likely that RUNX1 drives regulation of these processes. Key genes that regulate proliferation and apoptosis include *MYC* (Dang, 2012) and *BCL2* (Kale *et al.*, 2018). MLL-AF4 regulates both of these factors (Benito *et al.*, 2015; Robinson *et al.*, 2008), and as MLL-AF4 appears to increase TF binding affinity (Fig. 4.12 and Fig. 4.14), the overexpression of these genes may be partly driven by cooperative TF activity. I therefore investigated whether RUNX1 co-interacts with other TFs to drive *BCL2* and *MYC* expression.

The most frequent TFs to co-regulate the same genes as RUNX1 include core MLL-FP GRN TFs NF-YA, MYB, IKZF1 and MEF2C (Fig. 4.15A). Each of these factors also co-interact with Runx1 in EHT (Fig. 3.11D, p.85). Of these factors, MYB is predicted to regulate both *MYC* and *BCL2* along with RUNX1 (Fig. 4.15B-C). MAZ is a central TF that does not significantly co-interact with RUNX1, yet is predicted to regulate both gene targets. These factors are bound at both loci, the +1.7 Mb *MYC* superenhancer (Shi *et al.*, 2013), and a +129 kb *BCL2* enhancer (Fig. 4.15D). Other central factors predicted to regulate *BCL2*

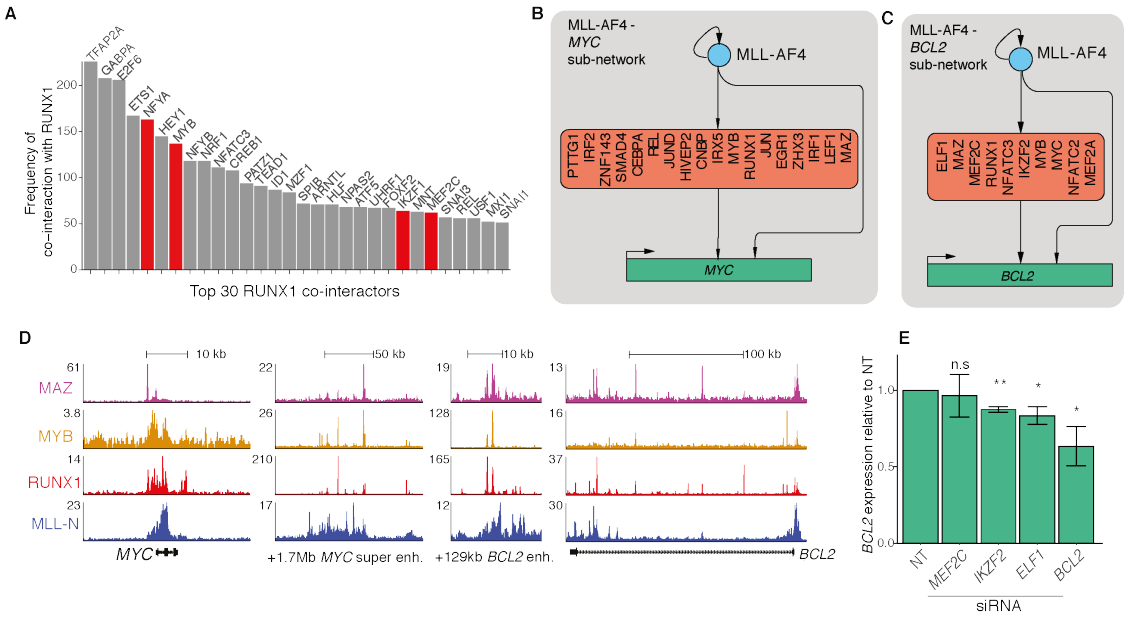


Figure 4.15: GRN predicted cooperative regulation of *MYC* and *BCL2*. (A) Frequency of TFs co-regulating gene targets with RUNX1. Core MLL-FP TFs highlighted in red. **(B, C)** Sub-networks of select interactions from MLL-AF4 and TFs regulating *MYC* (B) and *BCL2* (C). **(D)** ChIP-seq tracks for MAZ, MYB, RUNX1 and MLL-N at *MYC* and its super enhancer (1.7 Mb downstream of promoter), and *BCL2* and its enhancer (218 kb downstream of promoter). **(E)** qRT-PCR analysis probing *BCL2* pre-mRNA after 48 h siRNA treatment ($n = 3$). Expression normalized to *GAPDH*, and shown relative to NT control. Bars show mean with standard error; (*) $P < 0.05$; (**) $P < 0.01$. *Adapted from Harman* et al.*, 2021. Myb ChIP-seq data sourced from (Godfrey* et al.*, 2019).*

include MEF2C, IKZF2 and ELF1, of which *IKZF2* and *ELF1* KD lowers *BCL2* expression (Fig. 4.15E), validating these predicted interactions.

I performed KDs of *RUNX1*, *MYB* and *MAZ* alone or in combination with *RUNX1*. Perturbations reduced mRNA and protein levels with no interaction in co-treatments affecting siRNA efficiency (Fig. 4.16A-B). Due to stability of mature *BCL2* transcripts, intronic primers were used to probe *BCL2* pre-mRNA, as in Crump *et al.*, 2021. *RUNX1* KD alone was sufficient to reduce *BCL2* expression, but not *MYC* (Fig. 4.16C). *MAZ* KD significantly reduced *MYC* expression, in line with existing literature (Komatsu *et al.*, 1997), and borderline reduced *BCL2* expression ($P = 0.055$). *MYB* KD on its own showed no effect. However, combined *RUNX1* and *MYB* KD induced *MYC* and *BCL2* expression to trend towards

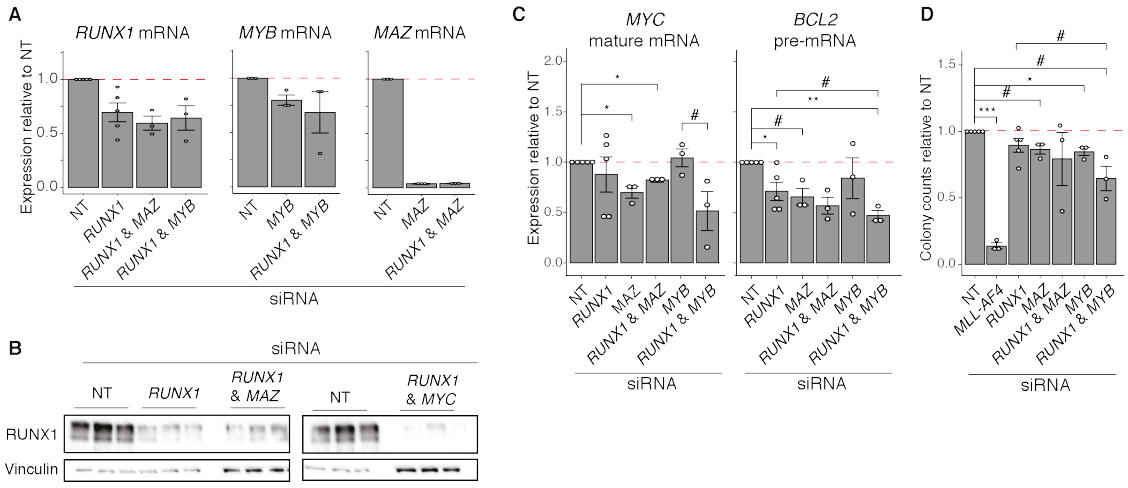


Figure 4.16: RUNX1 and MYB cooperate to drive *MYC* and *BCL2* expression. **(A)** qRT-PCR analysis for *RUNX1*, *MYB*, and *BCL2* after 96 hours' siRNA treatment ($n = 3$, $n = 5$ for *RUNX1* KD). Expression normalized to *GAPDH*, and shown relative to NT control. **(B)** Western blot for RUNX1 after 96 hours' siRNA treatment as indicated. Vinculin used as loading control. Three replicate experiments are shown. **(C)** qRT-PCR analysis probing mature *MYC* mRNA and *BCL2* pre-mRNA after 96 h siRNA treatment as indicated ($n = 3$, $n = 5$ for NT and *RUNX1* KD). Expression normalized to *GAPDH*, and shown relative to NT control. **(D)** Colony forming assay counts after 96 h siRNA treatment as indicated ($n = 3$, $n = 5$ for NT and *RUNX1* KD). Counts shown relative to NT control. Bars show mean with standard error; (#) $P < 0.1$; (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.001$. *Adapted from Harman* et al.*, 2021.*

downregulation, albeit not reaching significance (*MYC* P = 0.09; *BCL2* P = 0.063). *RUNX1* and *MAZ* KD showed no additional effect over individual perturbations. These results suggest that RUNX1 and MAZ function independently at these loci, while RUNX1 and MYB exhibit some level of co-regulation.

To see whether this cooperation extended to overall leukaemogenesis I used a CFU assay to assess growth following single or double TF KD (Fig. 4.16D). Note that the concentration of *RUNX1* siRNA used was lower than in Wilkinson *et al.*, 2013, so *RUNX1* KD RNA was incompletely perturbed (Fig. 4.16A-B) and resulted in minimal growth impact. *MAZ* KD on its own slightly reduced colony forming potential (P = 0.075), but additional *RUNX1* perturbation did not enhance this. *MYB* KD significantly impacted growth, and when combined with *RUNX1* KD the effect was increased though not reaching significance (P = 0.091). These assays further suggest that not only do RUNX1 and MYB co-interact to regulate *MYC*

and *BCL2*, but they also function to promote leukaemogenesis. Integrating these results together begins to interrogate the possible TF co-interaction at key loci, and provides a template for further study into gene regulation in leukaemia.

4.8 MLL-AF4 and cooperative TFs co-opt and drive normal fetal circuits.

I have established that MLL-AF4 and RUNX1 function cooperatively at many loci, and that TF cooperation drives key target expression. MLL-AF4 and RUNX1 regulatory logic also overlapped at B cell differentiation processes (Fig. 4.11B). *MLL* rearrangements in infant ALL (iALL) are thought to occur in fetal life (Ford *et al.*, 1993; Greaves, 2005, 2018; Jackson *et al.*, 2021), consistent with the observation that *MLL*r leukaemias are associated with overexpressed fetal programs (Rice *et al.*, 2021), and the MLL-AF4 GRN contains active FBM modules (Fig. 4.5). Thus, it is possible MLL-AF4 directly co-opts fetal TFs and pathways. To address whether normal fetal circuits are overactivated by the MLL-AF4 GRN, we collaborated with the Roy lab to generate single cell RNA+ATAC multiome (sc-multiome) data, profiling healthy CD34+ (stem and progenitor cells) and CD34- FBM and fetal liver (FL) cells from healthy donors. In addition, we profiled iALL and childhood ALL (chALL) blasts from patients, alongside xenograft models of MLL-AF4 transformed CD34+ FL cells generated by S. Rice using CRISPR induced translocation (Rice *et al.*, 2021) (Table 4.1). Samples were pooled in male-female pairs, and demultiplexed using a combination of *XIST* and Y chromosome expression and clustering of single cell variant calls (methods section 2.14.4, p.53). High quality, robustly deconvoluted cells were retained, and two samples (iALL-X and chALL-X) were discarded for poor cell capture (Table 4.1, Appendix A.5A-C).

The MLL-AF4 GRN is based on SEM cells, a B-cell ALL cell line. As such, I aimed to isolate HSPCs and B progenitors to capture the B lineage. To identify cell types, FBM/FL samples and MLL-AF4 ALL blasts were clustered, incorporating

Table 4.1: Patient and healthy donor samples used for single cell RNA+ATAC multiome analysis

ID	Population	Gender	Well	Cells pre-QC	Cells post-QC	Notes
FBM1.1	CD34- FBM	Male	1	3836	3146	
FBM2.1	CD34- FBM	Female	1	3810	2966	
FBM1.2	CD34+ FBM	Male	2	3748	3359	
FBM2.2	CD34+ FBM	Female	2	3123	2567	
FL1.1	CD34- FL	Male	3	3276	2736	
FL2.1	CD34- FL	Female	3	3165	2611	
FL1.2	CD34+ FL	Male	4	3129	2735	
FL2.2	CD34+ FL	Female	4	4998	4180	
iALL2	iALL blasts	Male	5	3901	3518	
iALL1	iALL blasts	Female	5	3653	2868	
iALL3	iALL blasts	Male	6	5435	4526	
chALL	chALL blasts	Male	6	5755	5160	
chALL-X	chALL blasts	Male	7	474	30	Fail
iALL-X	iALL blasts	Female	7	595	350	Fail
Xeno2-1	xenoALL blasts	Male	8	2856	2512	
Xeno3-3	xenoALL blasts	Female	8	3018	2724	
Xeno2-2	xenoALL blasts	Male	9	4994	4530	
Xeno3-2	xenoALL blasts	Female	9	6196	5752	
Failed deconvolution				8841		

both RNA expression and chromatin accessibility profiles (methods section 2.14.4, p.53, Hao *et al.*, 2021). Cells were annotated by label transfer using reference datasets (Jardine *et al.*, 2021; Popescu *et al.*, 2019; Roy *et al.*, 2021), and the B-cell lineage was manually curated using known transcriptional markers (Fig. 4.17A-B, Appendix A.6, Jardine *et al.*, 2021; Suo *et al.*, 2022).

Rather than focusing on individual clusters, I opted for a pseudotime approach to interrogate lineages, which allows for identification of expression patterns and quantitative changes in gene expression rather than cluster-dependent classification of marker genes. Slingshot trajectory analysis grouped cells into three lineages originating from HSCs (Defined by *HLF* expression, Lehnertz *et al.*, 2021), including erythropoiesis (*HBA1*), myelopoiesis/neutrophils (*MPO*, Aratani, 2018), and B lymphopoiesis (*PAX5*, Nutt *et al.*, 2001) (Fig. 4.17C-D). As a computational artifact the erythropoiesis lineage also picks up an eosinophil/basophil/mast cluster (Eo-Ba-Ma, *GATA2*, Ohmori *et al.*, 2015). Pseudotime ordering of these lineages organises cells into sequential clusters, from HSPCs through to mature populations

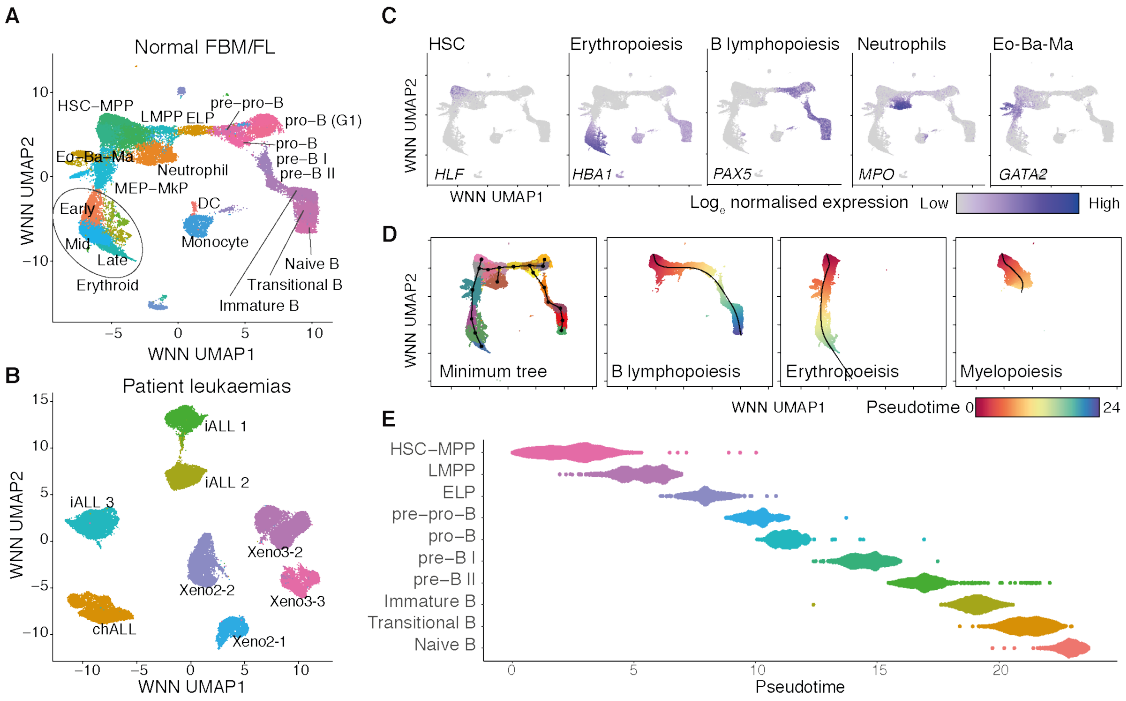


Figure 4.17: Sc-multiome trajectory analysis of FBM and FL cells recapitulates three fetal lineages. (A, B) UMAP dimensionality reduction and clustering performed using RNA+ATAC WNN graphs for normal FBM and FL cells (A), and patient and xenograft ALL blasts (B). Cluster identities annotated. **(C)** Expression of key lineage defining markers. **(D)** Lineage assignment through slingshot trajectory analysis, using HSC-MPP cluster as the lineage start. **(E)** B lymphopoiesis assigned cells ordered by slingshot pseudotime, stratified by assigned cluster as in A. *Analysis of the sc-multiome was performed by me in collaboration with A. Smith (scATAC-seq pre-processing).*

(Fig. 4.17E, Appendix A.15).

Differential analyses for expression changes over pseudotime was performed to generate a panel of B lineage associated genes (methods section 2.14.5, p.55), which were clustered into six B lymphopoiesis signatures (B1-6, Fig. 4.18A-B). As determined by single cell signature enrichment (AUcell), these map to HSC-MPP (B1), LMPP/ELP (B2), pre-pro-B/pro-B (B3), pre-B I and II (B4), immature/transitional B (B5) and Naïve B (B6) cells (Fig. 4.18C). To investigate how ALL gene expression compared with normal B lymphopoiesis expression, I probed the ALL samples with the B lymphopoiesis signatures. Interestingly only signatures pre-pro-B/pro-B (B3), and to a lesser extent LMPP/ELP (B2), were enriched in the iALL and chALL blasts (Fig. 4.18D). This suggests that ALL

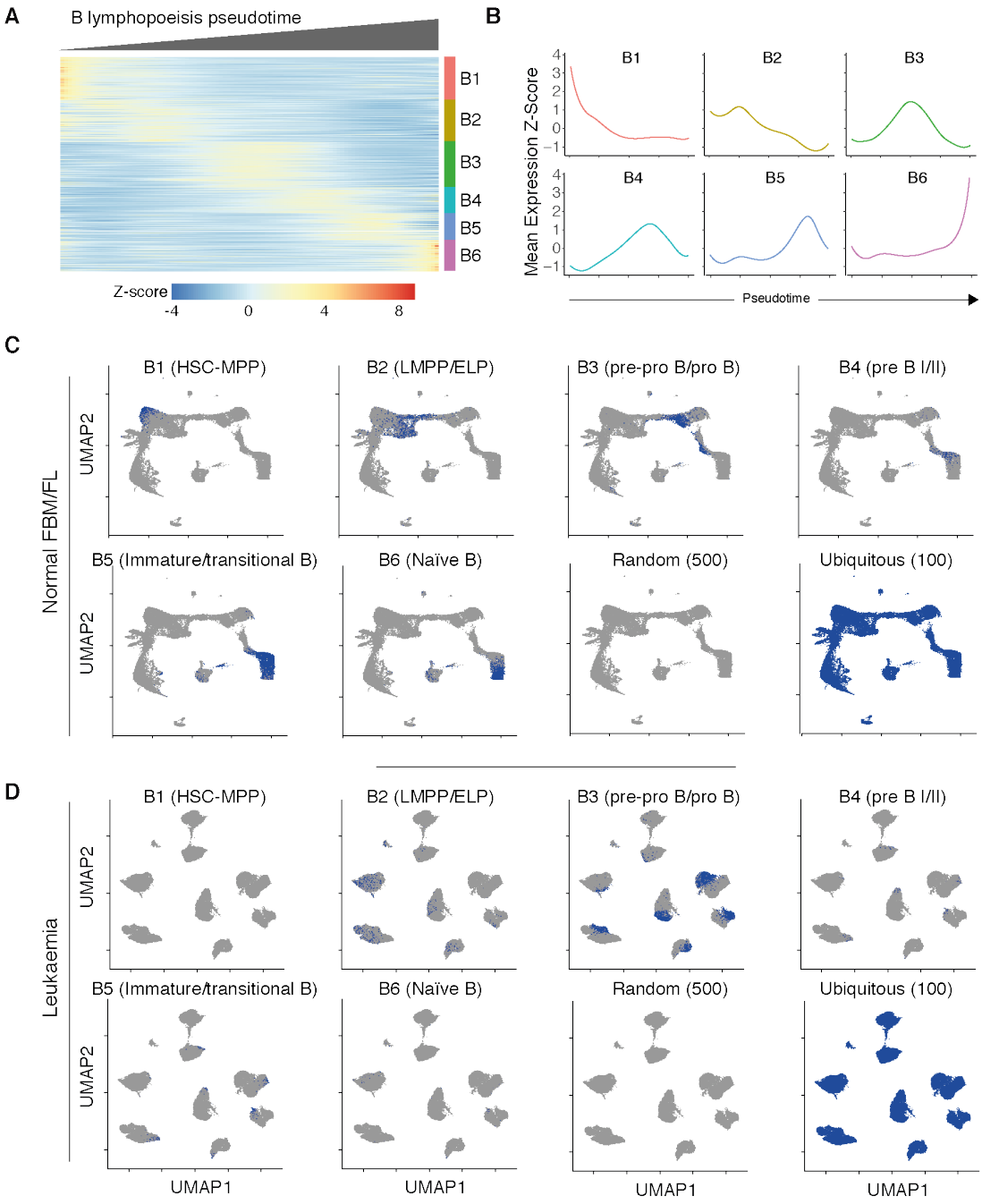


Figure 4.18: pre-pro-B/pro-B biased B-lymphopoiesis associated genes are enriched in MLL-AF4 ALL blasts. (A) Heatmap of z-score expression, smoothed over pseudotime into 100 bins, with genes as rows and cells as columns ordered by B lymphopoiesis pseudotime. *k*-means clusters (B lymphopoiesis signatures) shown on the side. **(B)** Average z-score expression, smoothed over pseudotime into 100 bins, split by clusters as in A. **(C-D)** FL and FBM (C) or ALL blast (D) UMAPs, with datapoints coloured blue for cells passing AUCell thresholds, marking enrichment for signatures as in A-B. The same thresholds were used for each signature between FBM/FL and ALL blasts. Random and ubiquitous gene sets are shown as controls.

blasts maintain transcriptional profiles from LMPP through to pro-B, but not more differentiated B lineage populations.

To determine whether MLL-AF4 co-opts B lymphopoiesis circuits I integrated the MLL-AF4 GRN with the B lymphopoiesis signatures. MLL-AF4 regulated genes were more significantly associated with B lymphopoiesis over non-MLL-AF4 regulated genes, but this was not observed for erythropoiesis or myelopoiesis (neutrophils) (Fig. 4.19A). This aligns with the hypothesis that MLL-AF4 drives B progenitor circuits. It is worth noting that while the patient expression data integration in section 4.3.1 identified a set of core nodes common across leukaemias, this was based on binary transcription and not levels of expression, and it also identified an ALL-specific program. In integrating the MLL-AF4 GRN with B lymphopoiesis signatures this analysis captures this ALL-specific program. Using the SEM *MLL-AF4* KD logFC data to assign activating and repressing interactions (as in Fig. 4.12), MLL-AF4 activated genes showed much greater B lymphopoiesis association than MLL-AF4 repressed genes (Fig. 4.19A). Strikingly, the GRN interfaces primarily with the B3 (pre-pro-B/pro-B) signature, and upon *MLL-AF4* KD this signature, unlike others, is almost entirely downregulated (Fig. 4.19B-C). Distinguishing MLL-AF4 direct and indirect binding reveals that most directly bound genes within each signature are downregulated upon KD, whereas indirectly regulated signatures are upregulated except for B3 (Fig. 4.19C). This can be interpreted as MLL-AF4 strongly activating a pre-pro-B/pro-B gene signature, both in the direct and indirect GRN, yet the indirect GRN may drive repression of stem and mature signatures.

A number of B lymphopoiesis associated genes are regulated by an MLL-AF4:RUNX1 driven FFL (Fig. 4.19D). Several FFL targets are associated with B lymphopoiesis, and are strongly regulated by MLL-AF4 (Fig. 4.19D-E). These include the key MLL-AF4 target *BCL2*, which shows a B2 expression pattern (ELP/LMPP). *LEF1* is a pre-pro-B/pro-B expressed gene that is a target of MLL-AF4 (Kerry *et al.*, 2017) and has a known role in driving pro-B proliferation as a part

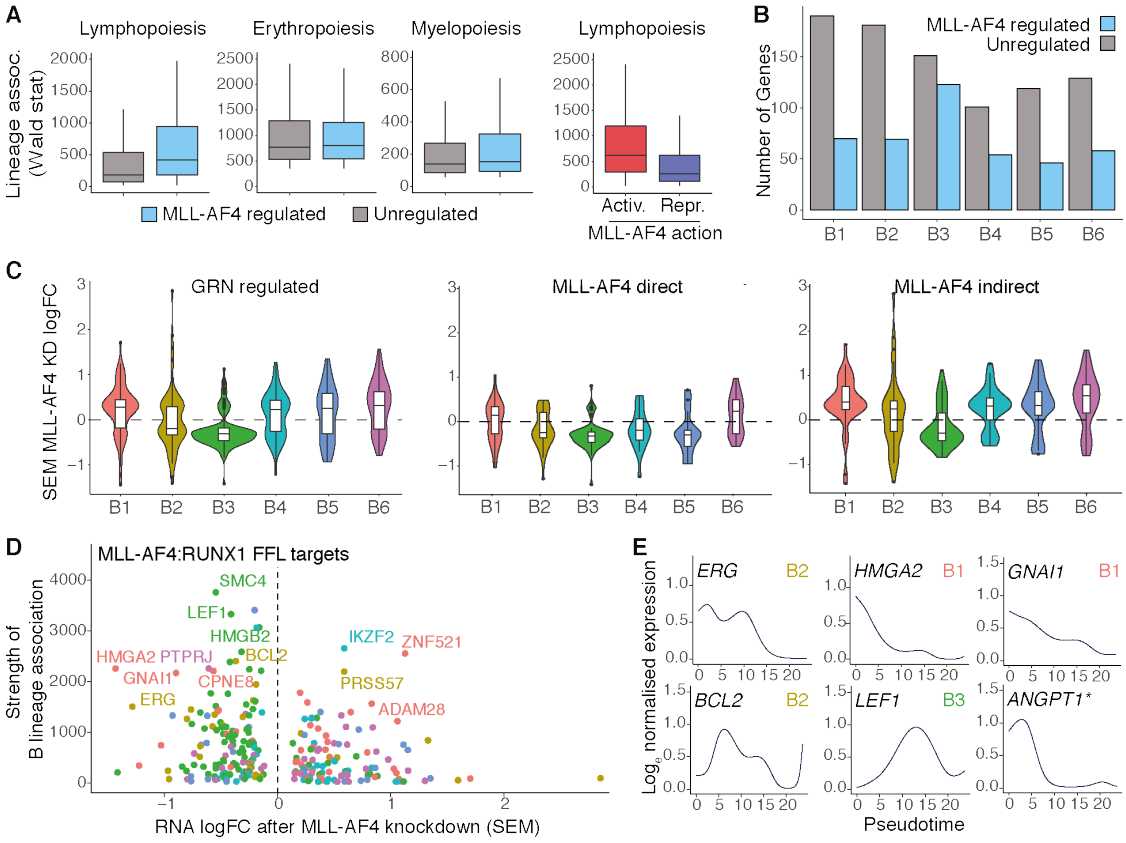


Figure 4.19: The MLL-AF4 GRN drives activation of pre-pro-B/pro-B biased genes, and may repress mature B-cell circuits. **(A)** Distribution of TradeSeq Wald statistics of lineage association for genes expressed in different lineages, split by MLL-AF4 GRN regulation status (left) or whether MLL-AF4 is activating or repressing, as determined in Fig. 4.12 (right). **(B)** Number of genes associated with each B lymphopoiesis signature, split by MLL-AF4 GRN regulation status. **(C)** Distribution of SEM RNA logFC following 96 hours' *MLL-AF4* KD across B lymphopoiesis signatures at all GRN regulated genes, MLL-AF4 directly, and indirectly regulated genes. **(D)** Relationship between SEM RNA logFC following 96 hours' *MLL-AF4* KD and TradeSeq Wald statistic of lineage association for B lymphopoiesis. MLL-AF4:RUNX1 FFL motif targets shown. Genes coloured according to B lymphopoiesis signatures. **(E)** TradeSeq smoothed expression plots for select genes in D, that are associated with B lymphopoiesis pseudotime. * Indicates genes were instead detected by pairwise differential expression between HSC and pro-B cells.

of WNT signalling (Reya *et al.*, 2000). *Lef1* was also highlighted in Maf:Lef1:EMT FFLs in mouse EHT (Fig. 3.16). One interesting hypothesis is that MLL-AF4 drives the reactivation of stem and progenitor programs, as seen with the overexpression of *PROM1* (Godfrey *et al.*, 2021; Mak *et al.*, 2012; O'Byrne *et al.*, 2019). In line with this, we find *HMGA2* and *GNAI1* matching a HSC-MPP signature (B1), and *ANGPT1*, which is not part of a signature but was instead identified by pairwise

differential expression analysis between HSC-MPP and pro-B clusters. This could reflect reactivation or forced maintenance of these loci.

One factor of interest is *ERG*, expressed in an ELP/LMPP biased signature (B2), which is expressed until pro-B and downregulated in pre-B cells. Pseudo-bulked accessibility profiles of the *ERG* locus in ALL blasts match pre-pro-B and pro-B cells, and the gene is bound by spreading MLL-AF4 and RUNX1 in SEM cells. This suggests MLL-AF4 and cooperative TFs bind to the accessible landscape in pro-B cells and prevents inactivation of the locus (Fig. 4.20A). There are two *ERG* transcripts commonly expressed in haematopoietic cells, *ERG2* (NM_004449) and *ERG3* (NM_182918) (Bohne *et al.*, 2009). Interestingly, MLL-AF4 and RUNX1 are only bound at the TSS of the *ERG2* transcript, and only *ERG2* transcripts are expressed (Fig. 4.20B), suggesting the MLL-AF4 drives a shift in *ERG* isoforms, though the consequence of this is not clear. *ERG* expression is required for V-to-DJ recombination and progression from a pro-B to pre-B state (Søndergaard *et al.*, 2019), yet *ERG* overexpression in mouse haematopoiesis also results in an expansion of pro-B cells and a partial pro-B to pre-B block (Tsuzuki *et al.*, 2011). As such, ERG may be required for V-to-DJ recombination but must be downregulated thereafter, hence forced expression could contribute to an ALL B differentiation block. In fact, ERG ChIP-seq shows binding at a higher proportion of the Naïve B signature (B6) genes compared to other signatures (Fig. 4.20C), suggesting downstream ERG interaction with mature B programs. Altogether, these analyses highlight the capacity for MLL-AF4, and cooperating TFs, to drive normal fetal circuits, reactivate stem and progenitor TFs, and potentially to repress mature B-cell programs, leading to a block in differentiation.

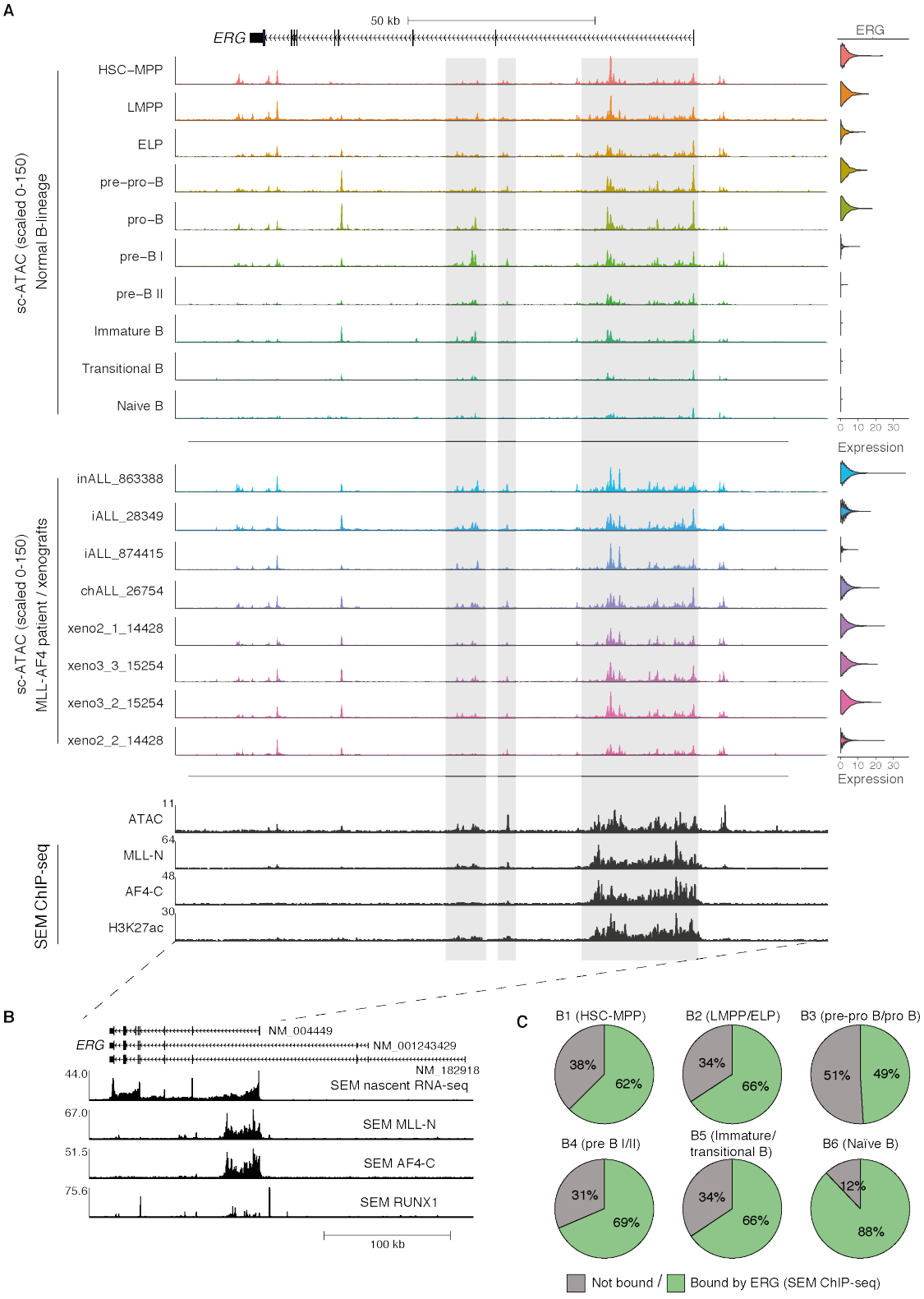


Figure 4.20: MLL-AF4 and RUNX1 regulate a specific isoform of the fetal B-progenitor TF, *ERG*. *Legend continued on next page.*

Figure 4.20: *Legend continued from previous page.* **(A)** Panel of ATAC-seq and ChIP-seq tracks at the *ERG* locus. scATAC-seq tracks are pseudobulked by cluster identity (FL/FBM) or sample identity (ALL blasts). Bulk ATAC-seq and ChIP-seq from SEM cells shown at the bottom. *ERG* expression shown on the right. **(B)** Zoomed out tracks showing three *ERG* isoforms, and SEM nascent RNA-seq and ChIP-seq. **(C)** Proportion of B lymphopoiesis signatures bound by ERG, based on ChIP-seq peaks. *ERG ChIP-seq data was sourced (Godfrey* et al.*, 2019).*

4.9 Conclusions and discussion

The central hypothesis of this chapter, based on the observation that *MLL-AF4* KD results in gene expression changes disconnected from direct MLL-AF4 binding, is that activated TFs are an essential component of *MLL*r networks. MLL-AF4 overexpresses a number of highly connected TFs, which may directly cooperate with MLL-AF4 to modulate gene expression, or act as an intermediate factor to regulate a wider network. To ask what the role of these activated TFs are in the wider MLL-AF4 network, I established GRN models describing the regulatory logic of MLL-AF4, the primary driver of leukaemic transformation (Andersson *et al.*, 2015; Bardini *et al.*, 2011; Bardini *et al.*, 2010), and RUNX1, a critical central TF.

By integrating the MLL-AF4 GRN with patient expression data, I was able to organise nodes into modules (section 4.3.1, Fig. 4.5). This identified a number of B progenitor nodes that are either maintained or overexpressed in ALL blasts (Patient module 1-3), as well as a number of genes that are uniquely activated in leukaemia (Patient module 4). This was an early suggestion that the MLL-AF4 network co-opts B progenitor circuits. This analysis also identified a cell line specific module (Patient module 5), which reflects an immortalised cell line phenotype. This module is unsurprising, considering the extensive transcriptional adaptations that occur in cell lines (Lopes-Ramos *et al.*, 2017), but highlights a problem with performing network analyses in cell line models. Perturbation studies are not feasible in donor patient samples, but cell line specific nodes can be removed through this integration with patient expression data. The most interesting observation was that many modules showed commonality across both AML and ALL samples (Patient modules 1, 2, 4), and that these were enriched for highly central TFs (section 4.3.2, Fig. 4.6). This analysis suggested the existence of a core network across both AML and ALL, and further validation in THP-1 cells confirmed common network interactions between MLL-AF4 and MLL-AF9 fusion proteins. This is an exciting observation, as this central network may anchor MLL-AF4 behaviour in both myeloid and lymphoid lineages and explain how *MLL*r leukaemias can

undergo lineage switching (Dorantes-Acosta *et al.*, 2012; Gardner *et al.*, 2016). While MLL-FP regulates similar central TFs across AML and ALL, the same is not true for RUNX1, which demonstrates context-specific differences. This leads to an interesting possibility, where MLL-AF4 activates a common panel of TFs, but these TFs can enact context-specific regulation.

Centrality analysis of the MLL-AF4 GRN, using degree and stress centralities, highlighted a number of interesting TFs (section 4.4.2). An important metric is the ratio of degree (connectivity) and stress centralities, which has been suggested to represent key nodes connecting distinct network modules (Joy *et al.*, 2005; Koschützki *et al.*, 2008). As such, we can infer that a high degree/stress ratio represents network paths with low robustness that may be susceptible to perturbation. RUNX1 and MYB are two TFs with a very high degree/stress ratio, and are both critically essential in MLL-AF4 and MLL-AF9 leukaemias. Additional central factors include ELF1, MAZ, MYC and NF-YA, among others. MAZ, despite high connectivity, had a low degree/stress ratio which was reflected in a lack of essentiality across multiple CRISPR screens. Despite this, MAZ was revealed to be a significant activator of *MYC*, in line with its discovery as a *MYC* associated zinc finger (Komatsu *et al.*, 1997), and to a lesser degree *BCL2*. NF-YA is a particularly interesting, and little explored target of MLL-AF4. By essentiality screens *NFYA* perturbation had little overall effect, and like MAZ it had a poor degree/stress ratio. Despite this, *MLL*r cell lines had overall greater essentiality scores for *NFYA* targeting in the DepMap database over non-*MLL*r blood cancers. A recent pre-print study has suggested a novel role in leukaemia, where NF-YA associates with, and directs the binding of Menin (Soto-Feliciano *et al.*, 2022). As Menin is a key MLL-FP complex component, this may direct binding of MLL-AF4 to non-uCpG sites. Other central factors of note were CEBPA and MED13L, which were not only highly connected, but showed much higher essentiality in *MLL*r leukaemias over non-*MLL*r. CEBPA also held a very high degree/stress ratio. CEBPA is commonly mutated in AML and thought to be involved in *MLL*r leukaemia initiation (Roe *et al.*,

2014). By contrast, MED13L has had little study in leukaemia, positioning it as a novel essential protein specific to *MLL*r cancers. The mediator complex, of which MED13L is a subunit of, acts as an intermediate complex associating TF binding with pol II activity, stabilises enhancer-promoter loops, and is a characteristic of super enhancers (Allen *et al.*, 2015; Whyte *et al.*, 2013). MLL-AF4 spreading is associated with MED1 (Kerry *et al.*, 2017), and MED26 interacts components of the SEC (Takahashi *et al.*, 2011), suggesting a significant interplay between MLL-FP and the mediator complex in regulating transcription.

Taking RUNX1 as an example central TF, my analysis led to a thorough investigation of the interplay between MLL-AF4 and RUNX1 (section 4.6). The role of TFs in the MLL-AF4 network can be broadly categorised into two types: TFs acting as an intermediate regulatory factor, as in the case of TF cascades, or through TF cooperativity forming FFL circuits. As such, MLL-AF4 and RUNX1 regulatory logic was organised into FFL and cascade motifs, and used to explore how MLL-AF4 could impact transcription beyond its direct binding to chromatin. This highlighted a number of MLL-AF4:RUNX1 driven cascade motifs, including an MLL-AF4:RUNX1:*CASP9* repressive chain that may promote cell survival. Interestingly, there were a large number of MLL-AF4:RUNX1 driven FFL motifs, and while RUNX1 can both activate and repress target genes (as seen with cascade motifs), in the context of MLL-AF4 binding RUNX1 regulatory logic is instead biased towards activation. It was also found that MLL-AF4 binding across chromatin was correlated with RUNX1 (section 4.6.2, Fig. 4.14), and appeared to promote TF:DNA binding affinity. Together, this strongly suggests that MLL-AF4 co-opts overexpressed TFs to further drive gene expression by cooperative regulation, and may promote the establishment of novel GRN interactions by encouraging the binding of TFs to DNA binding motifs (Fig. 4.21). MLL-AF4 also appears to alter TF activities, as seen where RUNX1 is biased towards activation at MLL-AF4 targets (Fig. 4.21). This TF cooperation was exemplified at *BCL2* and *MYC* where RUNX1 cooperates with

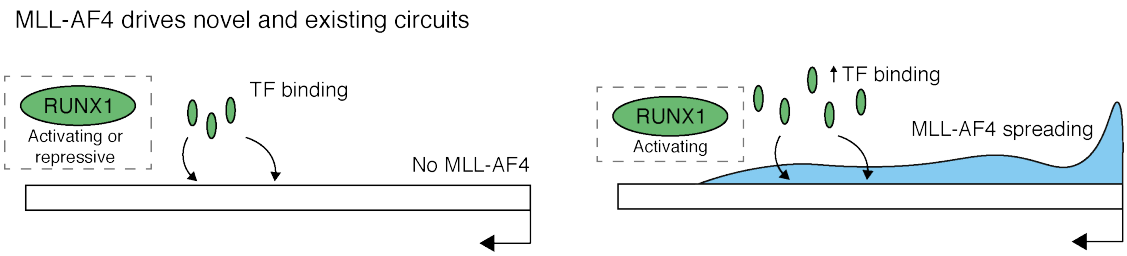


Figure 4.21: Summary model of how MLL-AF4 binding may alter TF activity. A model illustrating how MLL-AF4 binding may alter TF activity within the GRN model. MLL-AF4 spreading was shown to be associated with an increased frequency of TF peaks, which is most likely driven through increased affinity for DNA binding motifs. This increased affinity synergises with MLL-AF4, by encouraging binding of TFs, and may allow for novel regulatory interactions. Further, RUNX1 activity at non-MLL-AF4 sites shows activation and repression, whereas at MLL-AF4 bound genes it is biased towards activation.

MYB, in the context of MLL-AF4 binding, to enhance activation of these genes, leading to increased cell survival and proliferation (section 4.7, Fig. 4.22).

It was also noted that the MLL-AF4 network may maintain, or overexpress fetal circuits. Applying the MLL-AF4 GRN to B progenitor expression signatures highlighted significant interactions throughout B lymphopoiesis (section 4.8). In particular, the MLL-AF4 GRN appears to strongly activate a pre-pro-B/pro-B signature and repress stem and mature B-cell signatures. However, several TFs preferentially expressed in HSCs were regulated by MLL-AF4 in cooperation with RUNX1, including *HMGA2* and *ANGPT1*, which suggests a reactivation of HSC pathways (Fig. 4.23). *HMGA2* is a known target of MLL-AF4 that promotes leukaemic growth (Eguchi-Ishimae *et al.*, 2014; Guenther *et al.*, 2008), and has a role in genome wide remodelling of chromatin accessibility at AT rich sites (Vignali *et al.*, 2020). *HMGA2* by our analysis is typically downregulated as cells differentiate, and this maintained expression likely promotes global gene expression. *ANGPT1* (Angiopoietin-1) is also biased towards HSCs, where Angiopoietin-1 is secreted by HSCs to maintain their vascular niche in bone marrow (Zhou *et al.*, 2015). *ANGPT1* was previously found to be regulated by MLL-AF4 in SEM cells (Castro *et al.*, 2010), where it was noted that *ANGPT1* perturbation resulted in cell cycle arrest and reduced splenic infiltration in transplantation experiments. This

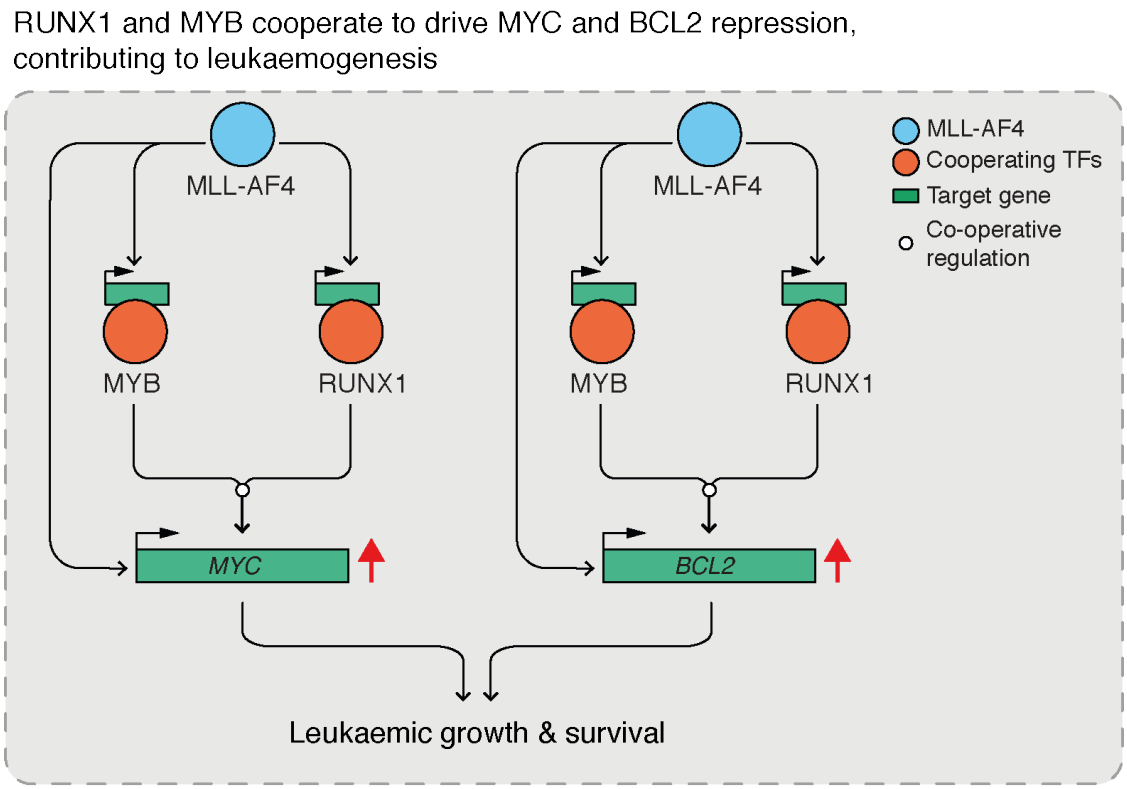


Figure 4.22: Summary model of RUNX1 and MYB cooperation to drive *MYC* and *BCL2* expression. A model of MLL-AF4-driven regulation of *MYC* and *BCL2*. RUNX1 and MYB cooperate to drive gene expression of both *MYC* and *BCL2*, which are known to functionally cooperate to drive leukaemic growth and survival (Bissonnette *et al.*, 1992; Fanidi *et al.*, 1992). *Figure is adapted from Harman* et al., *2021.*

highlights the utility of this analysis for identifying reactivation of early stem cell circuits in MLL-AF4 leukaemia, though the additional role of RUNX1 in regulating these targets in leukaemia has not previously been explored. *ERG* is a specific MLL-AF4:RUNX1 target highly regulated by MLL-AF4, that is normally expressed up to pro-B cells before repression in pre-B cells (Fig. 4.23). *ERG* is critical for V-to-DJ recombination, and promotes pro-B to pre-B differentiation (Søndergaard *et al.*, 2019), but this is incongruous with mouse experiments noting an expansion of pro-B cells and a partial pro-B to pre-B block with *ERG* overexpression (Tsuzuki *et al.*, 2011). One hypothesis that explains this is that ERG may be required for V-to-DJ recombination but must be downregulated thereafter for B progenitor differentiation. Therefore, forced expression by MLL-AF4 may contribute to an ALL B differentiation block (Fig. 4.23), though further study is required to determine

the validity of this hypothesis.

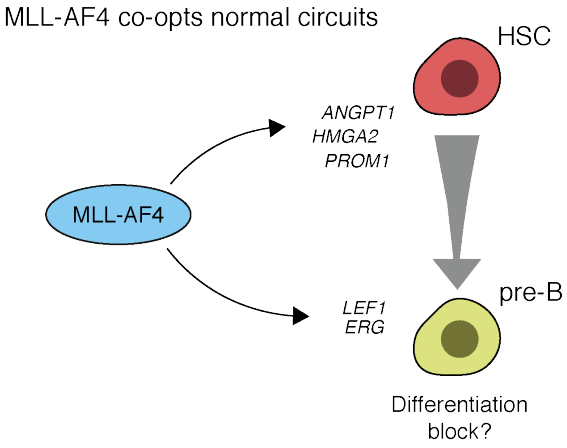


Figure 4.23: Summary model describing MLL-AF4 activation of developmental circuits. A model illustrating MLL-AF4 reactivation of early HSC-expressing genes, or maintenance of genes expressed in pro-B or pre-B cells. In doing so, MLL-AF4 can be considered to co-opt and force maintained activation of these circuits, which may contribute to the block in *MLL*r ALL differentiation.

5

A regulatory network driving MLL-AF9 GMP leukaemogenesis

Contents

5.1 Introduction and aims 152
5.2 Experimental design to characterise MLL-AF9 GMP leukaemogenesis 155
5.3 Gene expression and chromatin accessibility adapts over time, while MLL-AF9 binding is stable. 157
5.4 Generating a GRN model of MLL-AF9 GMP leukaemogenesis 160
5.5 Sub-networks can describe TFs driving expression modules 162
5.6 *Myc* upregulation occurs at week 3, after activation of multiple predicted regulators 164
5.7 Myb and MLL-AF9 are predicted to co-interact with a range of TFs to drive leukaemogenesis 166
5.8 Myb and Runx1 may cooperate to drive expression of key regulatory hubs. 168
5.9 Myb perturbation may cooperate with Runx1 to impact leukaemogenesis and dysregulate essential targets. 170
5.10 Conclusions and discussion 174

5.1 Introduction and aims

In chapters 3 and 4, I established and analysed GRN models describing aspects of normal and aberrant haematopoiesis. This involved a model of EHT, the process underlying the embryonic origin of HSCs, and a model of MLL-AF4 leukaemia. The MLL-AF4 GRN represented a steady state system, where the leukaemia has already progressed significantly. For this next chapter I have attempted to model early stages of leukaemogenesis, alongside conditional perturbation of the critical TF Myb in the presence of MLL-AF9, using a system previously established by I. Lau (Lau, 2022). This system involves the transduction of mouse granulocyte-monocyte progenitors (GMPs) with human MLL-AF9, and expansion of these cells in MethoCult. MLL-AF9 is known to activate *Myb*, which is then required for leukaemia maintenance (Zuber *et al.*, 2011a). MLL-AF9 also upregulates the micro-RNA (miRNA) *miR-196b* (Popovic *et al.*, 2009), which as a miRNA results in downstream gene repression (Huntzinger *et al.*, 2011). Previous work by I. Lau generated an engineered mouse model where two *miR-196b* miRNA response elements (MREs) were inserted into the 3' UTR of *Myb* (Lau, 2022), resulting in a conditional circuit where upon MLL-AF9 induction of *miR-196b*, *Myb* is targeted for repression (Fig. 5.1). In this chapter I have established a GRN model using

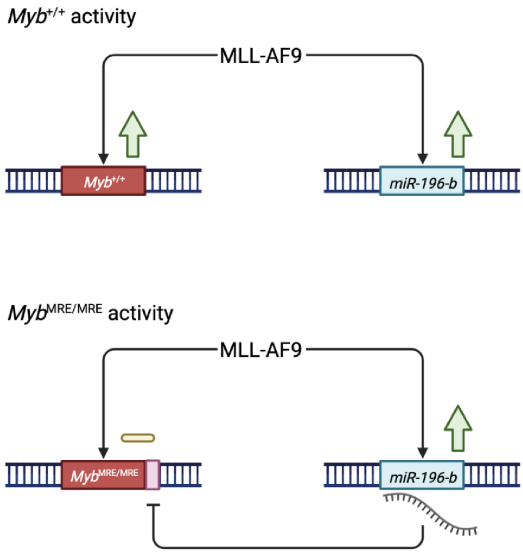


Figure 5.1: Schematic of *Myb*$^{\mathrm{MRE/MRE}}$ engineered circuit. Schematic illustrating *Myb*$^{+/+}$ and *Myb*$^{\mathrm{MRE/MRE}}$ genotypes, and the action of MLL-AF9 in these contexts. *Myb*$^{\mathrm{MRE/MRE}}$ includes a *miR-196b* response element in the 3' UTR of *Myb*, resulting *miR-196b* repression of *Myb*. As MLL-AF9 drives both *miR-196b* and *Myb* expression, this acts to intrinsically repress *Myb* activity during leukaemogenesis. *Myb*$^{MRE/MRE}$ *mouse models designed and generated by I. Lau (Lau, 2022). Figure created with BioRender.com.*

both $Myb^{+/+}$ and $Myb^{\mathrm{MRE/MRE}}$ mice, combined with MLL-AF9 transduction of GMPs, to describe early transcriptional adaptations to *MLL-AF9* expression. As this $Myb^{\mathrm{MRE/MRE}}$ conditional model was constructed to perturb *Myb*, I have focused the analysis of this chapter on the role of Myb in MLL-AF9 leukaemia.

The GRN models in chapters 3 and 4 were constructed using different methodologies, with advantages and disadvantages to each approach. The EHT model involved a thorough dissection of gene expression and chromatin accessibility patterns, integrated through motif detection. While this established a broad set of potential regulators, it lacked physical evidence of TF binding. The MLL-AF4 GRN instead relied on ChIP-seq data to identify regulatory interactions, and used perturbation data to focus the GRN on specific protein behaviours. It relied heavily on cell line models, which are poor representations of real patient biology. While this resulted in a robust and targeted GRN, the wider network was less reliable as the motif data was sourced from published deepCAGE data based on THP-1 cells, and there were not enough samples to perform correlation analyses. Going forward, to establish a GRN using this $Myb^{\mathrm{MRE/MRE}}$ model, I took lessons learned from chapters 3 and 4 to build a more complete network model. The aim was to establish a GRN that incorporates a time course, gene expression correlation, TF binding motifs, RNA-seq, ATAC-seq and ChIP-seq data, and perturbation data through the $Myb^{\mathrm{MRE/MRE}}$ mouse model. This will allow the generation of a robust network that can be used to probe the role of Myb as GMP regulatory networks adapt to the presence of MLL-AF9 over time, as a model of leukaemogenesis.

Aims

1. To characterise transcriptional and epigenetic adaptations to human *MLL-AF9* expression in GMPs.
2. To establish a GRN model of GMP leukaemogenesis over time.
3. To investigate the impact of *Myb* perturbation in the MLL-AF9 leukaemogenesis model.

The addition of MLL-AF9 to normal mouse GMPs allows the characterisation of MLL-AF9 leukaemogenesis and the construction of a GRN model that incorporates a time course experiment. This GRN model will highlight how regulatory circuits adapt over time to *MLL-AF9* expression. Further, by including a $Myb^{\mathrm{MRE/MRE}}$ mouse system, *Myb* perturbation data is incorporated in the model, allowing for specific questions on TF behaviour in the context of leukaemogenesis.

5.2 Experimental design to characterise MLL-AF9 GMP leukaemogenesis

To address how MLL-AF9 causes transcriptional and epigenetic adaptations over time, I established an experimental design to construct a GRN model incorporating multiple time points following *MLL-AF9* expression (Fig. 5.2A). GMPs were isolated from mouse bone marrow as Lin$^-$cKit$^+$CD41$^-$CD150$^-$CD16/32$^+$ (Fig. 5.2B), as in Pronk *et al.*, 2007. GMPs were retrovirally transduced with a construct containing human *MLL-AF9* expressed from the Murine Stem Cell Virus (MSCV) promoter and an *mCherry* fluorescent marker expressed from the 3-phosphoglycerate kinase

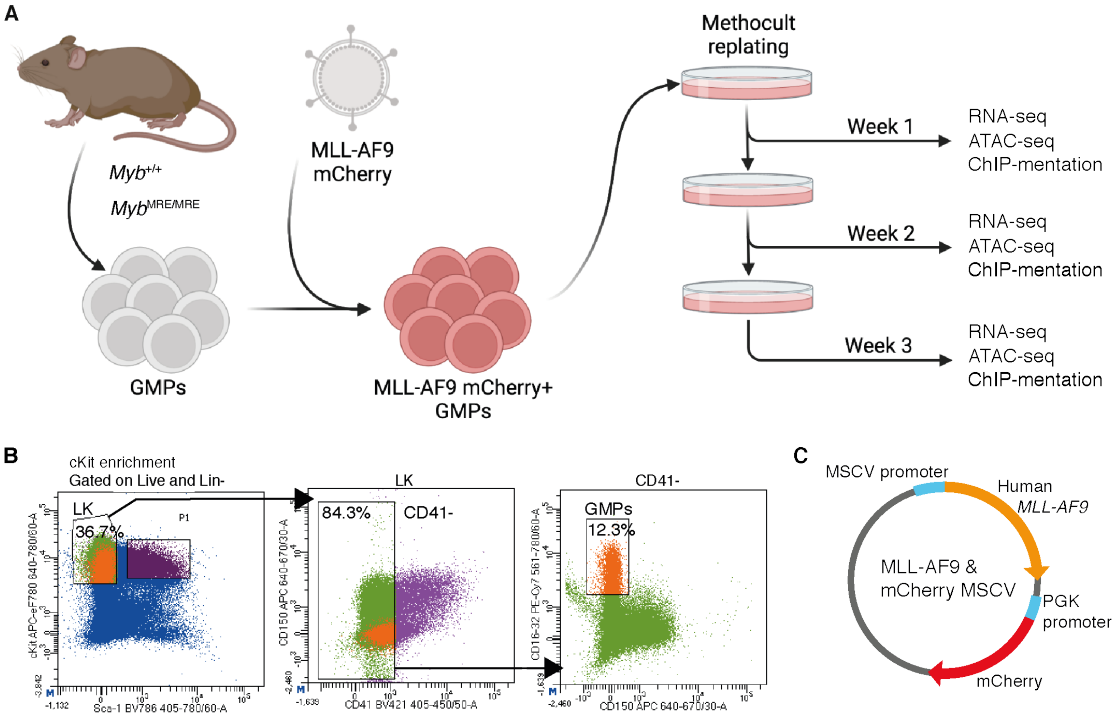


Figure 5.2: Experimental plan for MLL-AF9 transformation of GMPs and MethoCult expansion. (A) Schematic overview for plan to virally transduce mouse GMPs with a human MLL-AF9:mCherry construct and expand in MethoCult. Cells were replated in MethoCult each week. **(B)** Representative sort gates for isolation of GMPs from mouse bone marrow. c-Kit positive cells were isolated by magnetic bead purification before staining and flow cytometry. LK: Lin-cKit+. **(C)** Illustration of MSCV MLL-AF9:mCherry plasmid, with human *MLL-AF9* and *mCherry* constitutively active under the MSCV and PGK promoters, respectively. *Experimental design adapted from I. Lau, and MSCV MLL-AF9:mCherry construct generated by I. Lau (Lau, 2022). Panel A created with BioRender.com.*

(PGK) promoter (Fig. 5.2C). After 24 hours' incubation, mCherry+ GMPs were sorted directly into MethoCult. mCherry+ GMPs were expanded in MethoCult for three weeks with replating each week. Every week (W1-3) cells were collected for RNA-seq (5 x 10^5 cells), ATAC-seq (1 x 10^5 cells), and ChIPmentation (7 x 10^5 cells) (Fig. 5.2A). ChIPmentation was performed instead of ChIP-seq due to the limited cell numbers. This experiment was performed in parallel using GMPs sourced from both *Myb*$^{+/+}$ and *Myb*$^{MRE/MRE}$ mice, to allow comparisons between normal *Myb* activity and conditional *Myb* perturbation, and was performed in triplicate.

Maintenance of *MLL-AF9* expression after expansion in MethoCult was confirmed by RT-PCR, using primers targeting human MLL-AF9 (Fig. 5.3A), and the genotypes of *Myb*$^{+/+}$ and *Myb*$^{MRE/MRE}$ samples were validated (Fig. 5.3B). As noted by I. Lau (Lau, 2022), the *Myb*$^{MRE/MRE}$ genotype should confer reduced colony forming potential. In this experiment, by W3 *Myb*$^{+/+}$ samples showed considerably more colonies (Fig. 5.3C), although colony counting was only performed for two replicates.

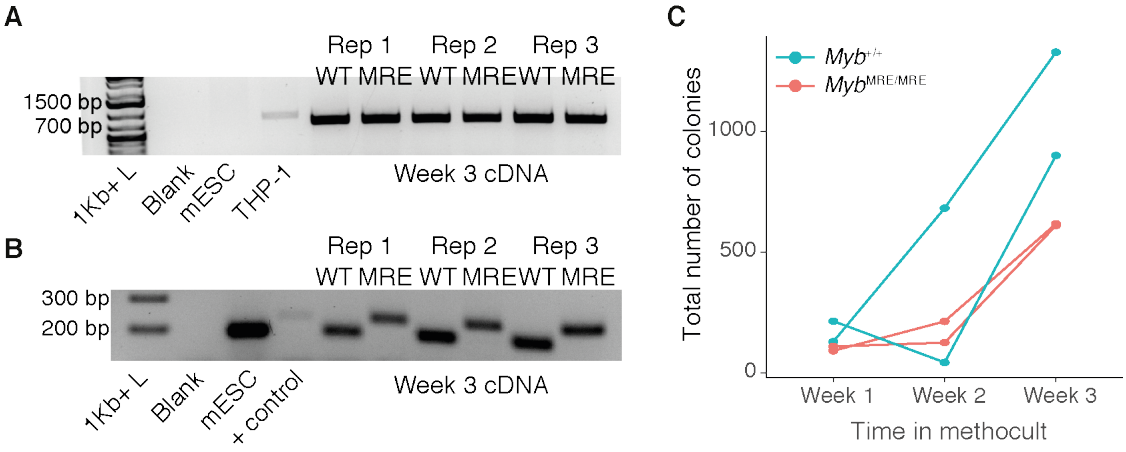


Figure 5.3: MLL-AF9 GMP replating validation. (A) PCR validation using cDNA from W3 MLL-AF9 GMPs showing MLL-AF9 expression (667 bp product). mESC and THP-1 cDNA used as negative and positive controls, respectively. O'GeneRuler 1Kb+ used for size ladder. **(B)** PCR validation using cDNA from W3 MLL-AF9 GMPs to confirm *Myb* genotypes. PCR products indicate *Myb*$^{+/+}$ at 186 bp and *Myb*$^{MRE/MRE}$ at 233 bp. O'GeneRuler 1Kb+ used for size ladder. **(C)** Total colony counts in MethoCult over three weeks of culture and replating. Week 1 counts were over one plate, while weeks 2 and 3 were counted over three plates per replicate. $n = 2$.

I performed the RNA-seq, ATAC-seq and ChIPmentation library preparation of the transformed cells at each time point. ChIPmentation was performed using antibodies raised against MLL-N, H3K27ac, and H3K4me1. Typical RNA Tapestation traces showed high quality RNA (Appendix A.7), with intact 18S and 28S ribosomal RNA peaks, and library prepped samples showed expected fragment sizes. ATAC-seq samples showed clear mono- and di-nucleosome fragments (Appendix A.7). After sequencing, RNA-seq and ATAC-seq samples showed a high number of PCR duplicates (RNA-seq: 37.15% - 67.22% PCR duplication, with mean of 59.75%. ATAC-seq: 28.27% - 52.67% PCR duplication, with mean of 40.2%) (Appendix A.8A-B). ChIPmentation samples had low read counts (6.09 x 10^6 - 1.35 x 10^7 read counts, with mean of 9.64 x 10^6) with few PCR duplicates (7.95% - 28.03% PCR duplication, with mean of 14.31%) (Appendix A.8C), suggesting that the ChIPmentation samples could be sequenced more deeply to increase complexity. After PCR deduplication, all samples were mapped successfully (Appendix A.8A-C). Overall, these quality control measures showed the data was of a reasonable quality for analysis.

5.3 Gene expression and chromatin accessibility adapts over time, while MLL-AF9 binding is stable.

To determine the MLL-AF9 binding pattern over time, MLL-N ChIPmentation was summarised over MLL-N peak calls (Fig. 5.4A). Interestingly, MLL-N signal is present at similar levels and distribution between W1 and W2, suggesting that MLL-AF9 binding profiles do not change over the time course. In comparison to $Myb^{+/+}$, the $Myb^{\mathrm{MRE/MRE}}$ samples, despite showing similar binding profiles, indicate reduced MLL-N binding. H3K27 acetylation shows a similar distribution of signal between W1 and W2, but with reduced levels at W2, for both genotypes. Note that the ChIPmentation samples are somewhat under-sequenced, and have not been reference normalised, so these observations are not fully robust. However, this

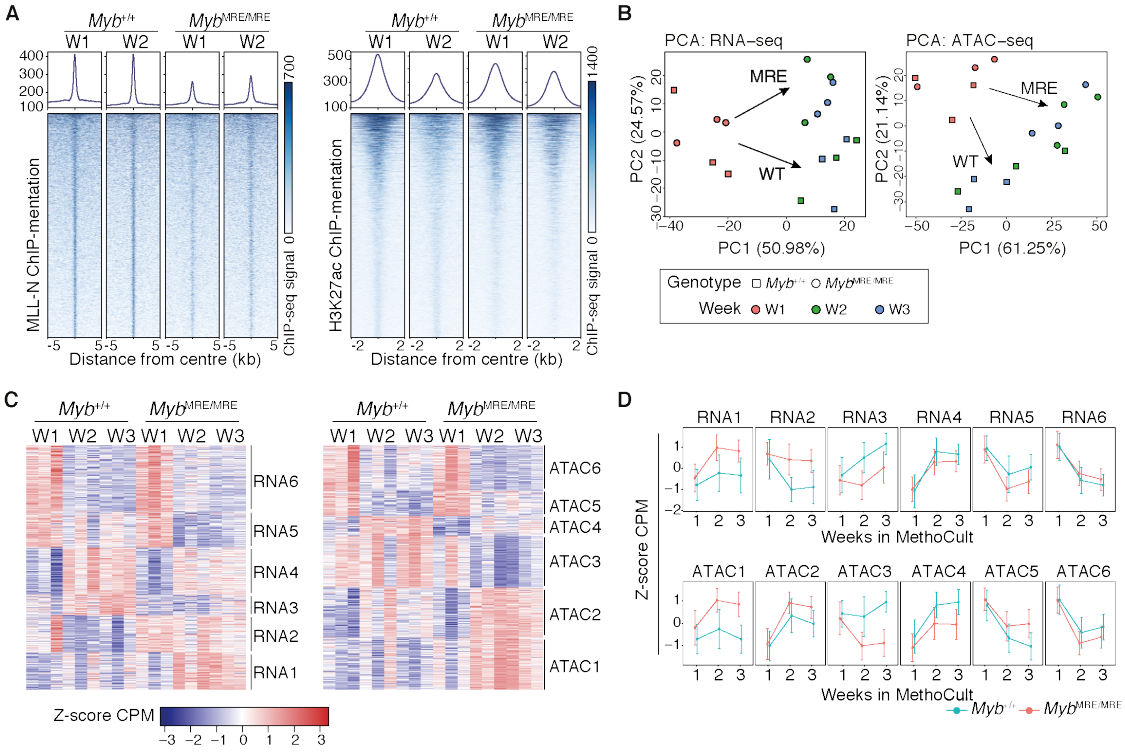


Figure 5.4: RNA and chromatin accessibility adapts from W1 to W2/3, yet MLL-N profiles are stable by W1. **(A)** MLL-N and H3K27ac ChIPmentation heatmaps in W1 and W2 MLL-AF9 transformed GMPs, at MLL-N peaks (left) and H3K27ac peaks (right), ordered by ChIPmentation signal. MLL-N peak calls were performed on combined W1 and W2 data. Metaplots showing average signal shown above. **(B)** PCA plots of RNA-seq (left) and ATAC-seq (right) samples, with ANOVA DEG and DAEs used as inputs. *Myb* genotypes represented by data point shape. Arrows indicate transition from W1 to W2 and W3 samples, for $Myb^{+/+}$ and $Myb^{\mathrm{MRE/MRE}}$. **(C)** Expression and accessibility heatmaps (z-score $\log_2$ CPM) for RNA-seq DEGs (left) and ATAC-seq DAEs (right). Genes and peaks grouped by k-means clustering into modules. **(D)** Average z-score $\log_2$ CPM of gene expression (top) and chromatin accessibility (bottom) by modules as in C. Bars indicate standard deviation of the mean.

analysis suggests that MLL-AF9 binding does not significantly change over time, though there may be differences between $Myb^{+/+}$ and $Myb^{\mathrm{MRE/MRE}}$ genotypes. In contrast, H3K27 acetylation appears to decrease over the time course, suggesting changes in the chromatin environment.

To address whether gene expression or chromatin accessibility profiles change during MethoCult expansion, despite stable MLL-AF9 binding, I performed differential analyses using an ANOVA-like method across all samples. This identified 4128 DEGs and 20527 DAEs. PCA dimensionality reduction in both RNA-seq and

ATAC-seq datasets found $Myb^{+/+}$ and $Myb^{MRE/MRE}$ clustered together at W1 (Fig. 5.4B), suggesting that the genotypes are similar at this time point. Interestingly, while $Myb^{+/+}$ and $Myb^{MRE/MRE}$ samples cluster similarly at W1 they diverge at W2 and W3. W2 and W3 samples are clustered similarly to each other, yet distinct from W1, highlighting considerable gene regulatory adaptations between W1 and W2. As transcriptional and chromatin accessibility changes occur from W1 to W2, this is accompanied by a divergence in behaviour with the conditional *Myb* perturbation.

Clustering the RNA-seq and ATAC-seq data into six modules highlights distinct patterns of regulation (Fig. 5.4C-D). Modules RNA1-2 and ATAC1-2 show higher expression and accessibility in $Myb^{MRE/MRE}$ over $Myb^{+/+}$, while RNA3/ATAC3 shows considerably higher expression in $Myb^{+/+}$, with unchanging or repressed activity in $Myb^{MRE/MRE}$ samples, suggesting that Myb repression has complex effects on the chromatin environment. RNA4 and RNA5 also shows greater expression in $Myb^{+/+}$, albeit to a lesser degree. RNA6 and ATAC6 shows less difference between *Myb* genotypes. Overall, RNA1, RNA3, and RNA4 show upregulation of gene expression over time, while RNA2, RNA5, and RNA6 show downregulation over time. As expected from the PCA analysis, the greatest magnitude of change occurs between W1 and W2. These modules highlight specific expression patterns, with RNA3 of particular interest as it highlights impaired gene activation under *Myb* perturbation. Further, paired with the observation that MLL-N binding profiles do not significantly change over time (Fig. 5.4A), this analysis suggests that most genes are not affected immediately. This delayed sensitivity could be explained by secondary effects of the network, where key regulators must be switched on before changes occur.

5.4 Generating a GRN model of MLL-AF9 GMP leukaemogenesis

To investigate the regulatory logic driving both the transcriptional adaptations occurring over time, as well as the differences between $Myb^{+/+}$ and $Myb^{\mathrm{MRE/MRE}}$ genotypes, I constructed a GRN model. Enhancer annotations performed in section 3.3, p.73 relied on annotating accessible elements to the nearest differentially expressed promoter, and additionally assumed that any accessible regions had enhancer potential. To improve upon this enhancer annotation method, I first filtered differential ATAC-seq peaks for overlapping H3K27ac ChIPmentation peak calls, then applied an iterative approach that allows for flexible annotations dependent on the gene density of the surrounding genome (Fig. 5.5A, Appendix A.10, methods section 2.15.4, p.58).

In brief, differentially accessible enhancers were associated with all expressed promoters within a 50 kb window. To seed potential E-P interactions, I defined enhancers with three or more promoter links as sufficiently annotated (resolved). Unresolved enhancers underwent iterative annotations by increasing the annotation window by 10 kb each step, until resolved. For peaks with no annotations in a 3 Mb window, the nearest expressed promoter was annotated. Using this approach, the majority of peaks are resolved within a 550 kb window (Fig. 5.5B), and most genes had fewer than 10 enhancers annotated (Fig. 5.5C).

As in section 3.3 p.73, to retain high confidence associations, E-P interactions were additionally filtered for greater than 0.4 correlation between enhancer ATAC-seq signal and promoter RNA expression (Fig. 5.5D). This method successfully allows annotating long-distance enhancers that span gene deserts, such as the *Myc* +1.7 Mb super enhancer (Shi *et al.*, 2013) (Appendix A.17), while limiting over-annotation in gene dense regions. E-P interactions are typically constrained by the boundaries of topologically associated domains (TADs, Beagan *et al.*, 2020). Using published TAD data (Wilson *et al.*, 2016; Xiang *et al.*, 2020) generated in

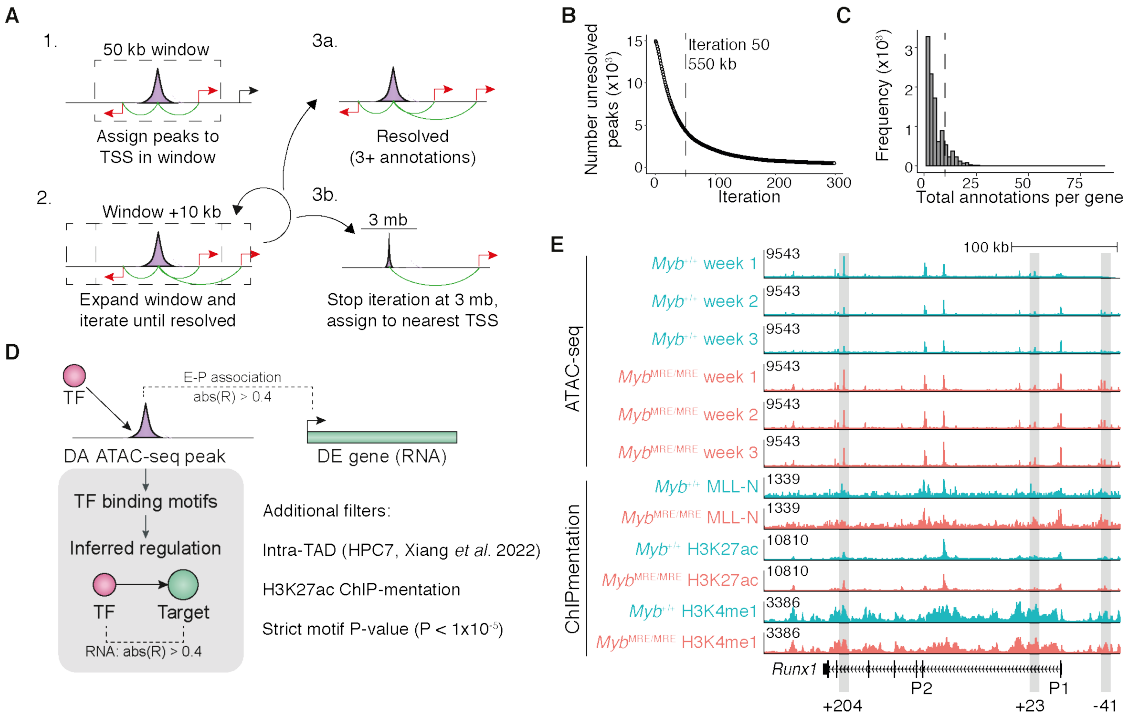


Figure 5.5: Constructing a GRN model of MLL-AF9 GMP transformation. **(A)** Schematic of iterative peak calling approach. Peaks were assigned to any TSS in a 50 kb window, with further annotations added with incrementally larger windows (+10 kb per step). This iteration is performed for each peak, until resolved (3+ annotations per peak), or the window reaches 3 Mb. **(B)** Number of unresolved peaks remaining at each step in the iterative peak calling method. Dashed line indicates the point of diminished returns, with majority of peaks resolved. **(C)** Number of peaks annotated to each gene promoter using the iterative peak annotation approach. Dashed line indicates 10 annotations per gene. **(D)** Schematic outlining approach to generate the MLL-AF9 GMP GRN. **(E)** ATAC-seq and ChIPmentation tracks at *Runx1*. GRN predicted differential enhancer elements highlighted in grey. ChIPmentation time points were combined into a single track for visualisation.

the HPC7 cell line (ESC derived HPC model, Pinto do O *et al.*, 1998), 76.2% of E-P interactions assigned were found within TADs. In this analysis I reasoned that HPC7s were a reasonable model of GMPs, as TAD boundaries are thought to vary little between cell types (Beagan *et al.*, 2020). E-P interactions that spanned TAD boundaries were removed, as these likely represent erroneous annotations. Overall, this approach successfully annotates long-distance enhancers, such as the *Myc* super enhancer, however without supplementary TAD data it over-calls distal enhancers.

To assemble the E-P interactions into a GRN model, as in section 3.3 p.73,

upstream regulators were inferred by the presence of TF binding motifs at enhancers and promoters (Fig. 5.5D). To improve confidence, motifs were filtered by P-value $< 1 \times 10^{-6}$, which still allows detection of the relatively short canonical Myb motif (5'-AACNG-3', Ogata *et al.*, 1994). Regulatory interactions were filtered for correlation greater than 0.4 (or less than -0.4) between expression of the TF regulator and predicted target gene. The MLL-N ChIPmentation data was used to infer MLL-AF9 binding. This method is exemplified at the *Runx1* locus, where the +23 and +204 enhancers were successfully identified, and are marked by H3K27ac and H3K4me1, as well as binding by MLL-AF9 (Fig. 5.5E). Together these data were integrated to establish an overall GRN model of the regulatory adaptations occurring after MLL-AF9 transduction.

5.5 Sub-networks can describe TFs driving expression modules

The GRN model is based on DEGs covering all time points, and includes gene expression data from both $Myb^{+/+}$ and $Myb^{MRE/MRE}$ samples. To determine the key drivers of this generalised GRN model, I calculated the most central nodes (Fig. 5.6A). As the key driver of the leukaemia, MLL-AF9 is the most connected node. Interestingly, despite the impact of *Myb* perturbation on MLL-AF9 leukaemia, Myb is not among the most central factors, at least within the overall network. While Myb is not the most connected, it may still regulate essential components of the leukaemic GRN. Similar to the MLL-AF4 GRN (Fig. 4.3B, p.110), Elf1, Egr1 and Mef2c are central factors in the MLL-AF9 GMP GRN, with Egr1 displaying the highest betweenness centrality (Fig. 5.6A).

As the GRN model describes both $Myb^{+/+}$ and $Myb^{MRE/MRE}$ gene expression, it is important to then decode the network to query specific questions. The RNA gene expression modules (RNA1-6, Fig. 5.4C-D) were used to decode the network by creating GRN sub-networks based on module genes, followed by recalculation

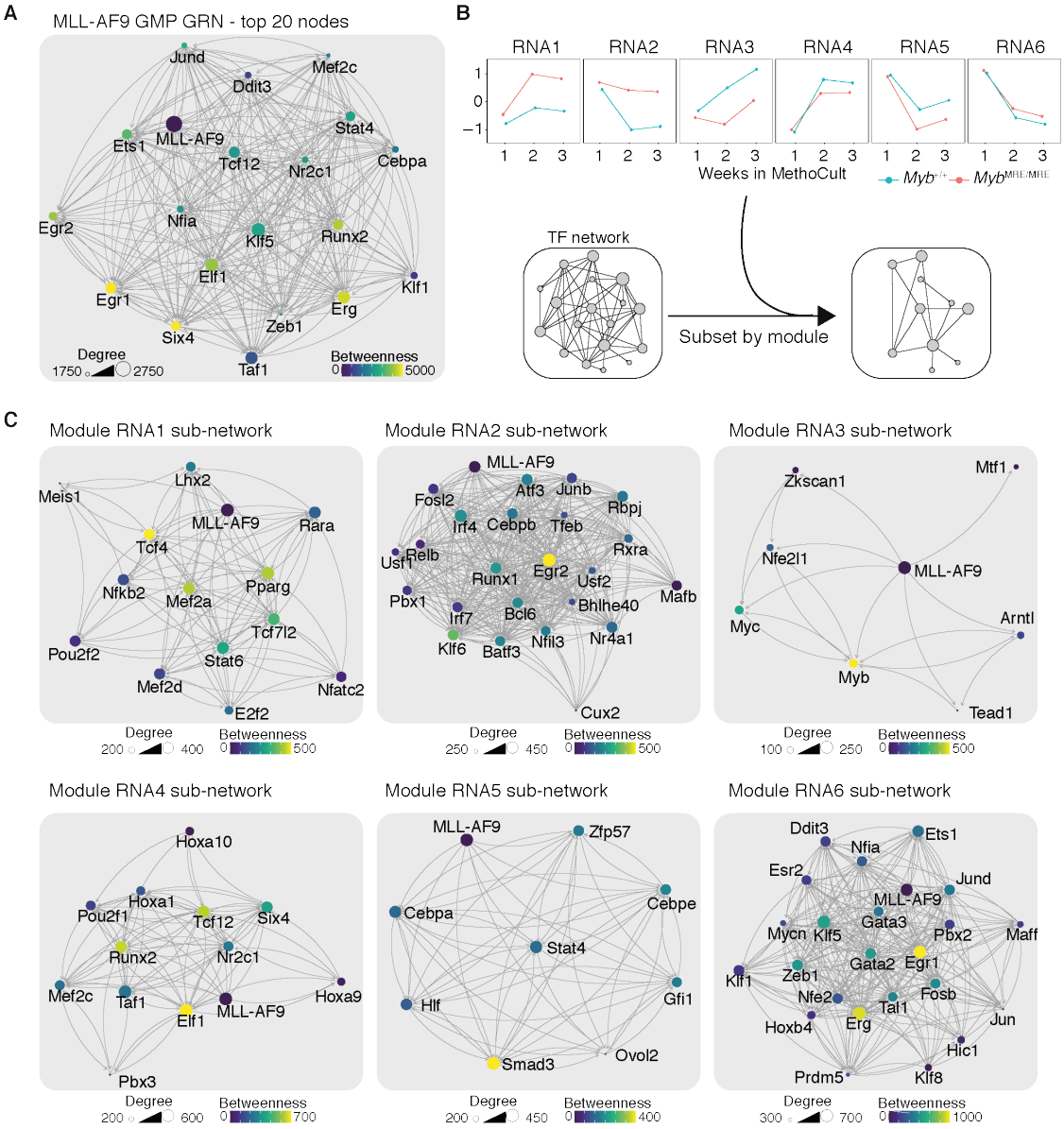


Figure 5.6: Generation and visualisation of MLL-AF9 GMP GRN sub-networks. (A) The top 20 genes of the MLL-AF9 GMP GRN by degree centrality. **(B)** Illustration of sub-network generation, by filtering the overall MLL-AF9 GMP GRN by RNA-seq module genes (RNA1-6), derived in Fig. 5.4C-D. **(C)** Sub-networks based on RNA modules, generated as in B. Sub-networks show all hub-node TFs expressed in each module. For all network visualisations, size represents degree centrality, while colour represents betweenness centrality. Lines indicate predicted interaction from protein to gene locus, with arrowheads pointing downstream.

of centrality measures (Fig. 5.6B). Module RNA3, encompassing TFs that display much higher expression in $Myb^{+/+}$ samples over $Myb^{MRE/MRE}$, contains both Myb and Myc, both of which display high degree and betweenness centrality (Fig. 5.6C). This aligns with the critical function of both these TFs in leukaemia, and suggests that they may be driving the $Myb^{MRE/MRE}$ phenotype. RNA4 genes show slightly increased expression in $Myb^{+/+}$ over $Myb^{MRE/MRE}$, and the TFs Runx2 and Elf1 show the highest degree and betweenness centrality within its sub-network. This module also contains Hoxa1, Hoxa9, and Hoxa10, suggesting the *Hoxa* cluster overall is repressed in $Myb^{MRE/MRE}$ samples. This analysis confirms that Myb is an important driver of transcription, most specifically within the RNA3 sub-network, and is associated with significant downregulation in the $Myb^{MRE/MRE}$ genotype. Similarly, TFs such as Runx2 and Elf1, as well as multiple hox genes, may drive lesser downregulation in the $Myb^{MRE/MRE}$ genotype, through the RNA4 sub-network.

5.6 *Myc* upregulation occurs at week 3, after activation of multiple predicted regulators

As discussed in section 4.7 p.133, *Myc* is a critical target in *MLL*r leukaemias that we found to be activated synergistically by both Runx1 and Myb (Fig. 4.16, p.135, Harman *et al.*, 2021). The clustering and sub-network analysis (section 5.5) placed *Myc* in module RNA3. However, while *Myc* does show slightly decreased expression in $Myb^{MRE/MRE}$ samples, it is not affected to the same extent as other RNA3 members, such as *Myb* (Fig. 5.7A). Interestingly, while the majority of transcriptional changes in this system occur between W1 and W2 (Fig. 5.4B-D), *Myc* expression is primarily upregulated from W2 to W3 (Fig. 5.7A).

To understand this delayed activation of *Myc*, I explored its predicted upstream regulators (Fig. 5.7B-C). *Myb* expression is reasonably correlated with *Myc*, though *Myb* perturbation in $Myb^{MRE/MRE}$ does not have a strong effect on *Myc* regulation. *Runx1* is instead anti-correlated with *Myc*, which disagrees with the MLL-AF4

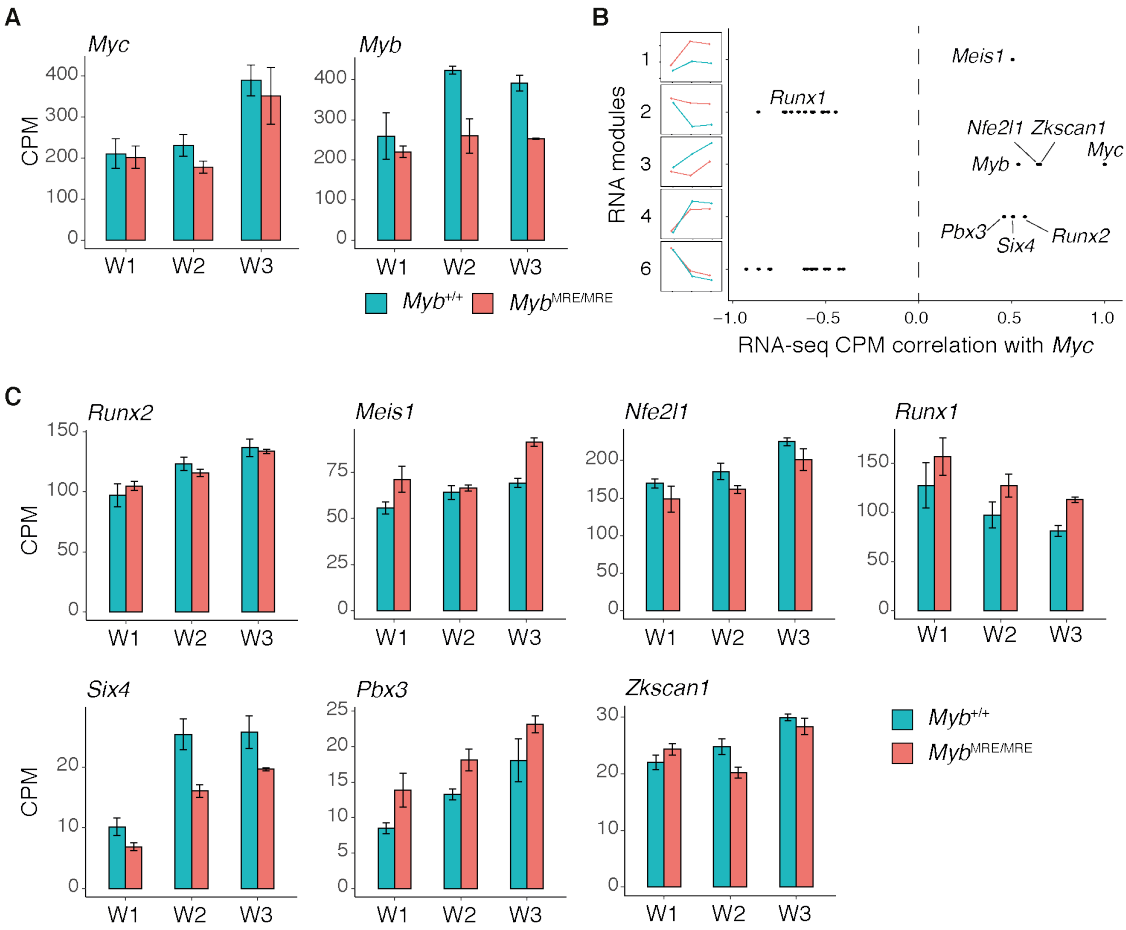


Figure 5.7: *Myc* is not fully expressed until week 3, potentially through additional TF regulators. (A) Gene expression (CPM) for *Myc* and *Myb*. Error bars represent standard error of the mean; $n = 3$. **(B)** RNA correlation between *Myc* and predicted *Myc* regulators, stratified by RNA-seq expression modules. Illustrative representation of modules shown along the y-axis. **(C)** Gene expression (CPM) for predicted *Myc* regulators positively correlated with *Myc* expression. Error bars represent standard error of the mean; $n = 3$.

GRN model (Fig. 4.16, p.135), and could suggest that Runx1 activity differs in the MLL-AF9 GMP context. In fact, *Runx1* deletion is known to impair lymphopoiesis but increase myeloid cell expansion (Growney *et al.*, 2005; Ichikawa *et al.*, 2004), and has been proposed to repress *Myc* in AML via a FOXC3/RUNX1/HDAC1/TLE3 repressor complex (Simeoni *et al.*, 2021). Multiple predicted regulators are positively correlated with *Myc* expression, including *Runx2*, *Meis1*, and *Nfe2l1* (Fig. 5.7B-C). However, none of these TFs strongly correlate with *Myc*, and they are not upregulated primarily from W2 to W3, but instead generally show continual

upregulation from W1 to W3. It is possible that multiple TFs must first be activated and could be required to cooperate before *Myc* upregulation occurs, contributing to this delayed gene expression response.

5.7 Myb and MLL-AF9 are predicted to co-interact with a range of TFs to drive leukaemogenesis

To investigate how Myb and MLL-AF9 may cooperate with other TFs, I applied the same analytical approach as in section 3.5, p.83 (methods section 2.15.6, p.60) to identify probable cases of TF co-interaction. As previously, I performed Fisher's exact tests for predicted targets across all TF pairs, and clustered odds ratio scores describing the likelihood that TF pairs co-interact more frequently than random. This resulted in six co-interaction clusters representing groups of TFs that may cooperate (CoInt 1-6, Fig. 5.8A). CoInt-1 contains Hox TFs and CoInt-4 contains Gfi1 and Meis1, and both clusters are enriched for general haematopoiesis and differentiation terms (Fig. 5.8A-B). CoInt-2 contains Myb, Runx1, and MLL-AF9 and is enriched for differentiation processes in both myeloid and lymphoid lineages (Fig. 5.8A-B). CoInt-5 contains several central factors, including Elf1 and Runx2, but shows little pathway enrichment, while CoInt-6 includes Smad3 and is correspondingly enriched for BMP signalling (Fig. 5.8A-B).

As Myb and MLL-AF9 are key factors in this system, I focused the analysis to predict their co-interaction partners (Fig. 5.8C). MLL-AF9 has widespread binding and regulates many targets, but still shows significant co-interaction with specific TFs. These include Hoxa9 and Hoxa10, Nfil3, Cebpb, and Cux2. For Myb co-interactions, among RNA3 module genes Arntl and Nfe2l1 are both predicted to co-interact with Myb. Arntl (known as Bmal1) is a core circadian rhythm TF that has been shown to be important *MLL*r AML progression (Puram *et al.*, 2016). It is possible that Arntl cooperates with Myb to affect downstream targets, as part

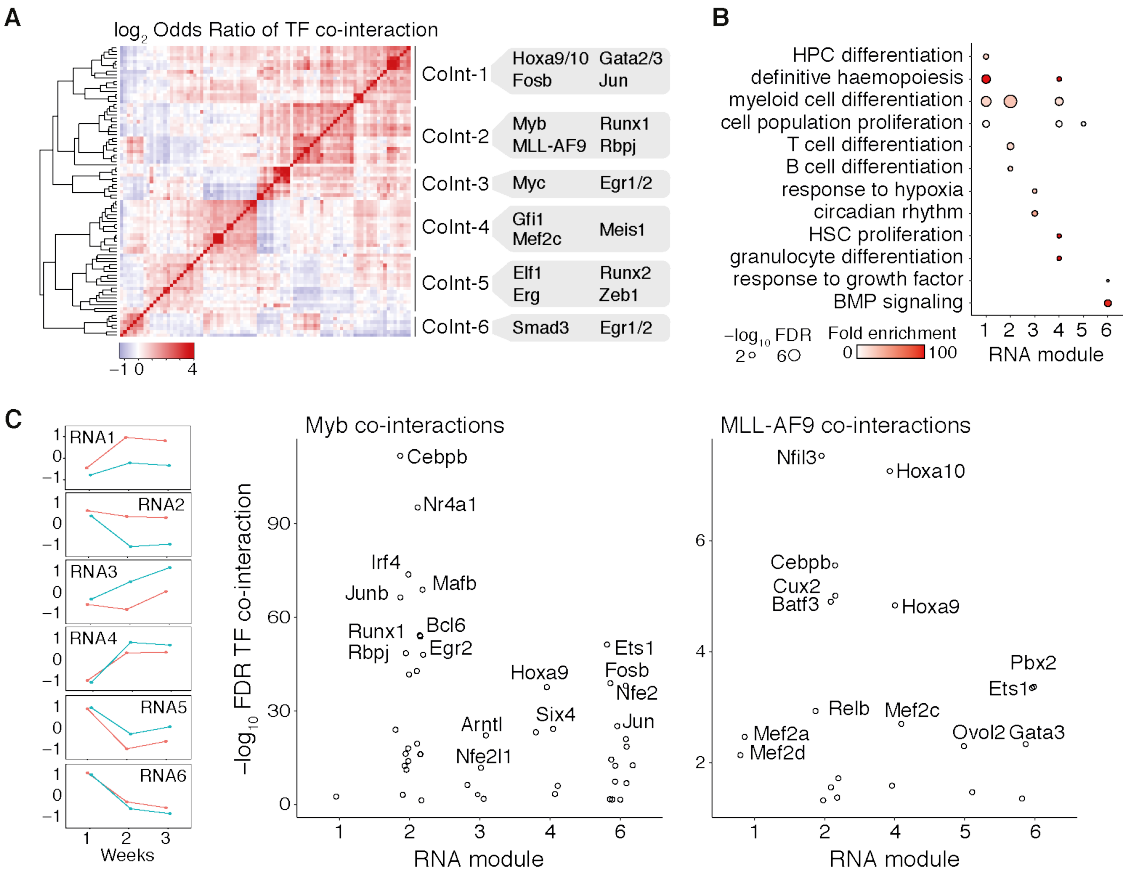


Figure 5.8: TF co-interaction differentiation, growth, and signalling pathways. **(A)** Hierarchical clustered heatmap of Fisher's exact test odds ratios comparing the overlap of TF-targets between pairs of TFs. Benjamini and Hochberg correction of P values was performed, and TFs not significant for any interaction were excluded. Clusters describe sets of TFs that commonly act at the same loci, suggesting cooperative regulation. Example TFs present in each cluster are given. **(B)** GO biological process enrichment for clusters CoInt 1-6, as in A. Only significant data points shown ($P \leq 0.05$). **(C)** Significance of TF co-interactions with Myb or MLL-AF9, as calculated in A, stratified by RNA modules (RNA1-6). RNA-seq expression modules shown on left for reference.

of the circadian rhythm. The most significant Myb co-interactors are expressed in RNA2, which shows an opposite expression pattern to Myb, and includes Cebpb, Nr4a1, and Irf4. Interestingly, the well-known MLL-FP target *Hoxa9* (Milne *et al.*, 2010) significantly co-interacts with both Myb and MLL-AF9. Despite a differing expression pattern, Runx1 still significantly co-interacts with Myb, suggesting that while Runx1 may have different activity in AML it could still cooperate with Myb. This preliminary analysis begins to explore potential TF cooperation that may drive key processes. Further investigation is necessary to establish key regulatory motifs.

5.8 Myb and Runx1 may cooperate to drive expression of key regulatory hubs.

As we have previously established, MYB and RUNX1 function cooperatively in the regulation of *Myc* in MLL-AF4 ALL (section 4.7, p.133), and Myb appears to significantly co-interact with Runx1 in the MLL-AF9 GMP GRN (Fig. 5.8C). I therefore explored potential cooperative regulation driven by these TFs. While Runx1 shows a different expression pattern to Myb (RNA2 and RNA3 respectively), and may act repressively in AML (Simeoni *et al.*, 2021), Runx1 is expressed throughout the time course and as such could still cooperate with Myb. To determine whether these TFs regulate similar targets I compared their regulatory logic. Both Myb and Runx1 preferentially target genes in expression modules RNA2, RNA3, RNA4 and RNA6, with few targets in RNA1 and RNA5 (Fig. 5.9A). In comparison, MLL-AF9 binds a high proportion of genes across all RNA modules. In addition to regulating similar gene classes, 48.5% of Myb targeted accessible elements are also predicted targets of Runx1 (Fig. 5.9B). This shows that alongside a high significance of co-interaction, Myb and Runx1 show similar regulatory patterns across expression modules and a high rate of overlap across enhancers and promoters.

As TFs are important regulators of *MLL*r leukaemias, I explored whether Runx1 and Myb co-regulate GRN regulatory hubs (Fig. 5.9C). Despite Runx1 and Myb co-interacting at 48.5% of Myb-targeted elements, of predicted Myb-regulated TFs (28 genes) 75% (21 genes) are also predicted to be regulated by Runx1. Focusing specifically on RNA3 genes, which displayed the greatest difference between $Myb^{+/+}$ and $Myb^{\mathrm{MRE/MRE}}$ genotypes, Myb and Runx1 appear to co-regulate *Myc*, *Nfe2l1*, and *Arntl* (Fig. 5.9D). As *Runx1* expression decreases over time while these targets are upregulated, it might be expected that Runx1 plays an inconsequential or repressive role at these sites. However, it is possible that Runx1 behaves differently at sites also bound by Myb, or that Runx1 supports Myb activity. It has been reported that Runx1 acts synergistically with Myb in neutrophil myelopoiesis in

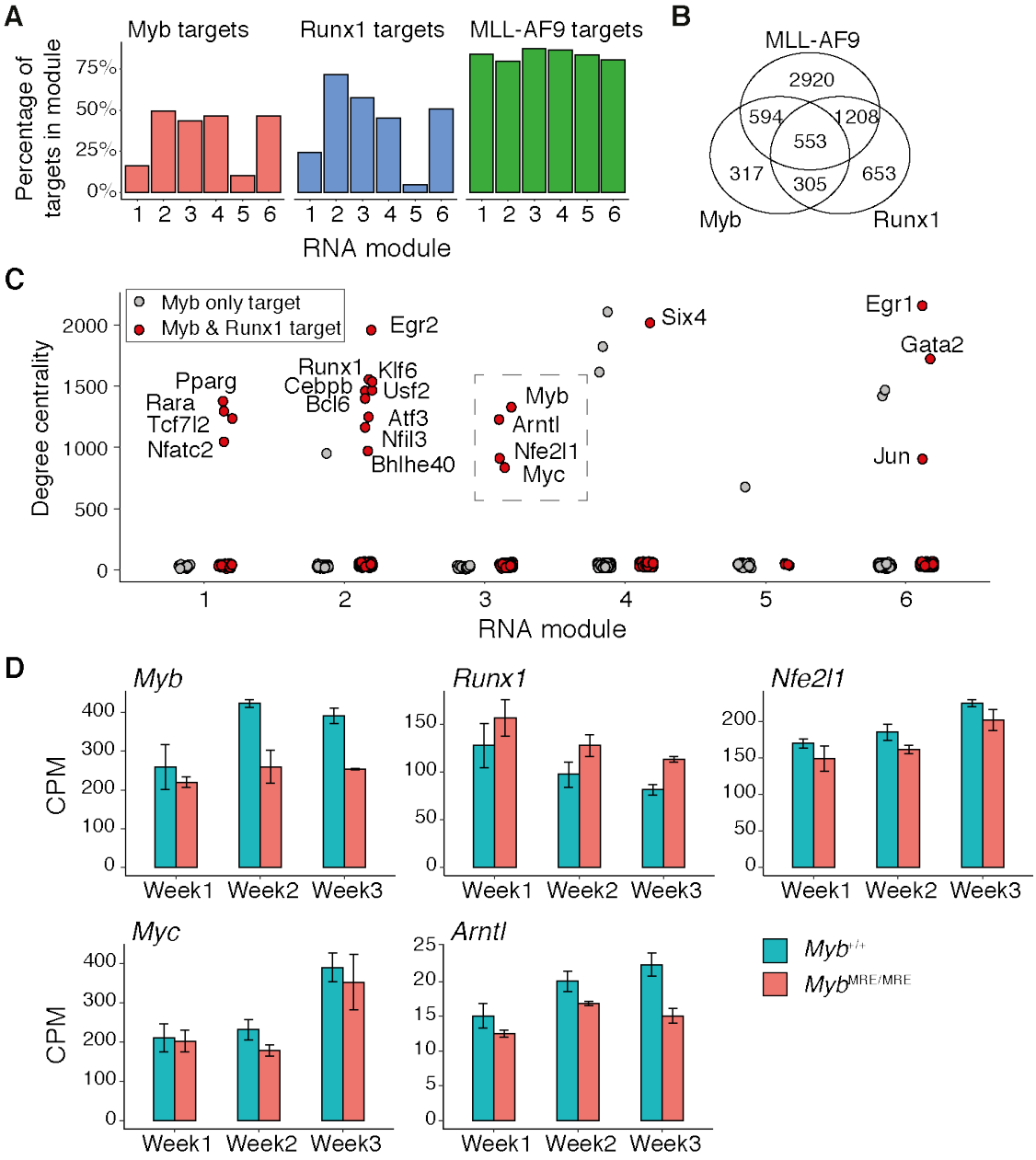


Figure 5.9: Myb and Runx1 co-interact at multiple loci. **(A)** Proportion of genes within each RNA-seq expression module with predicted targeting by Myb, Runx1, or MLL-AF9. **(B)** Overlap of nodes between predicted Myb, Runx1, and MLL-AF9 target accessible elements. **(C)** Stratification of degree centrality (MLL-AF9 GMP GRN) by RNA-seq expression modules. Data points coloured by whether they are targeted by Myb, or Myb and Runx1. Dashed box highlights RNA3 genes with predicted regulation from both Runx1 and Myb. **(D)** CPM expression for *Runx1* and RNA3 module genes targeted by both Myb and Runx1. Error bars represent standard error of the mean; $n = 3$.

zebrafish (Huang *et al.*, 2021) and so it remains possible, though as yet unexplored, that Runx1 positively regulates these targets along with Myb.

5.9 Myb perturbation may cooperate with Runx1 to impact leukaemogenesis and dysregulate essential targets.

It is possible that Runx1 promotes greater Myb regulatory potential. While we do not have *Runx1* perturbation data in this system to investigate this possibility, we can examine whether Myb:Runx1 co-regulated genes are more sensitive to *Myb* suppression. I performed a differential comparison between $Myb^{+/+}$ and $Myb^{\mathrm{MRE/MRE}}$ samples. In this comparison, I modelled time (Weeks 1-3) as a blocking factor, such that variation due to time in culture is removed (methods section 2.14.1, p.50). DEGs from this test therefore represented general differences in *Myb* genotype, rather than week-specific effects. This resulted in 1031 differentially expressed GRN nodes (Fig. 5.10A). Downregulated genes in this comparison (lower expression in $Myb^{\mathrm{MRE/MRE}}$) were primarily in expression modules RNA3, RNA4 and RNA5, whereas RNA1 and RNA2 contained upregulated genes. To interrogate positive Myb activity in the context of Runx1 predicted regulation, I explored logFC distributions in RNA3-5, and found that at Myb:Runx1 co-regulated targets, gene expression was reduced to a greater extent in $Myb^{\mathrm{MRE/MRE}}$ than genes predicted to be regulated by Myb without Runx1, though the effect was subtle (Fig. 5.10B). The difference in logFC was greater at RNA4-5 than RNA3. This effect was also seen in RNA1 and RNA2, where Myb:Runx1 predicted co-regulation confers greater upregulation. This analysis suggests that Myb may enact stronger regulation in the context of Runx1, though whether this is through direct cooperation with Runx1 needs to be explored.

To focus on the activity of Myb, I used the RNA-seq module RNA3, which shows the greatest difference in expression with *Myb* perturbation ($Myb^{\mathrm{MRE/MRE}}$). I took predicted Myb targets in the RNA3 expression module and stratified these by logFC

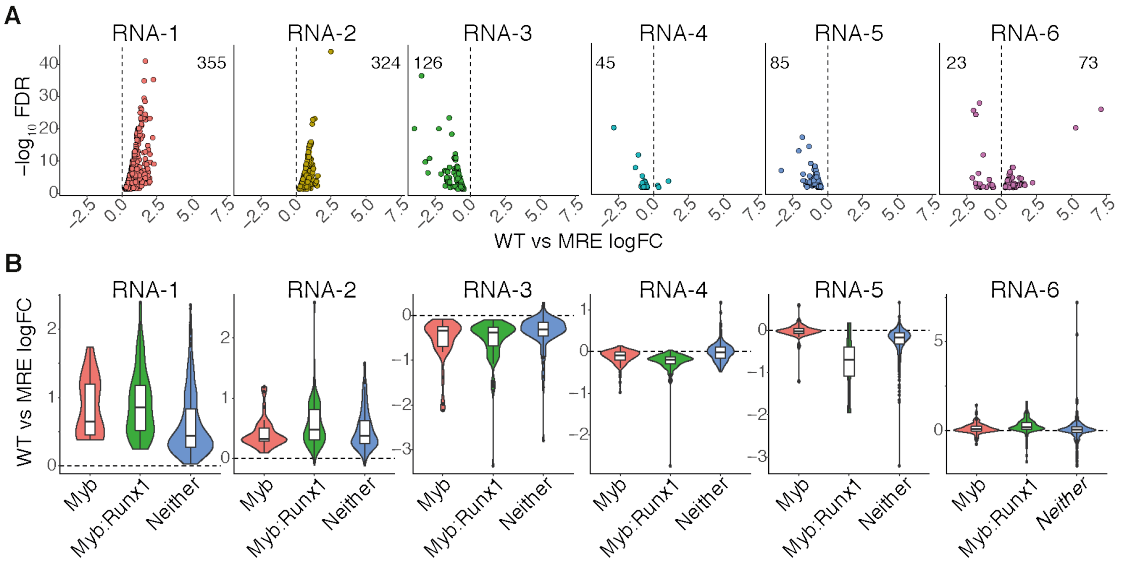


Figure 5.10: Myb and Myb:Runx1 target genes are more sensitive to *Myb* perturbation by the *Myb*$^{MRE/MRE}$ circuit. (A) Relationship between -$\log_{10}$ FDR and logFC (comparing *Myb*$^{+/+}$ and *Myb*$^{MRE/MRE}$; positive logFC means higher expression in *Myb*$^{MRE/MRE}$), in the MLL-AF9 GMP RNA-seq data. Results are stratified and coloured by RNA-seq expression modules. **(B)** Violin and boxplot summary of logFC values (comparing *Myb*$^{+/+}$ and *Myb*$^{MRE/MRE}$), stratified by RNA-seq expression modules. Data coloured by whether genes are targeted by Myb only, or Runx1 and Myb, or neither.

(comparing *Myb*$^{+/+}$ and *Myb*$^{MRE/MRE}$) and gene expression correlation with *Myb* (Fig. 5.11A). Two genes highly downregulated in *Myb*$^{MRE/MRE}$, and highly correlated with *Myb* expression (R > 0.7), were Hoxa11 and Bcat1. *Hoxa11* is predicted to be regulated by Myb and not Runx1, is downregulated in *Myb*$^{MRE/MRE}$ samples, and is highly correlated with *Myb* expression (Fig. 5.11A-B). *Bcat1*, which is predicted to be co-regulated by both Myb and Runx1, is moderately downregulated in *Myb*$^{MRE/MRE}$ samples and has very high *Myb* expression correlation (R > 0.9). A previous study has found *Bcat1* to be overexpressed in AML patients and associated with poor prognosis, as well as to drive DNA hypermethylation similar to IDH^{mut} AML (Raffel *et al.*, 2017). In contrast to the *Myb*$^{+/+}$ samples, in *Myb*$^{MRE/MRE}$ *Bcat1* expression remains the same across all weeks (Fig. 5.11B), suggesting that MLL-AF9 alone does not drive expression over MethoCult expansion and that this overexpression is dependent on Myb. This preliminary exploration begins to identify potential mechanisms by which Myb promotes leukaemogenesis, and in the case of *Bcat1* may indicate a mechanism independent of MLL-AF9.

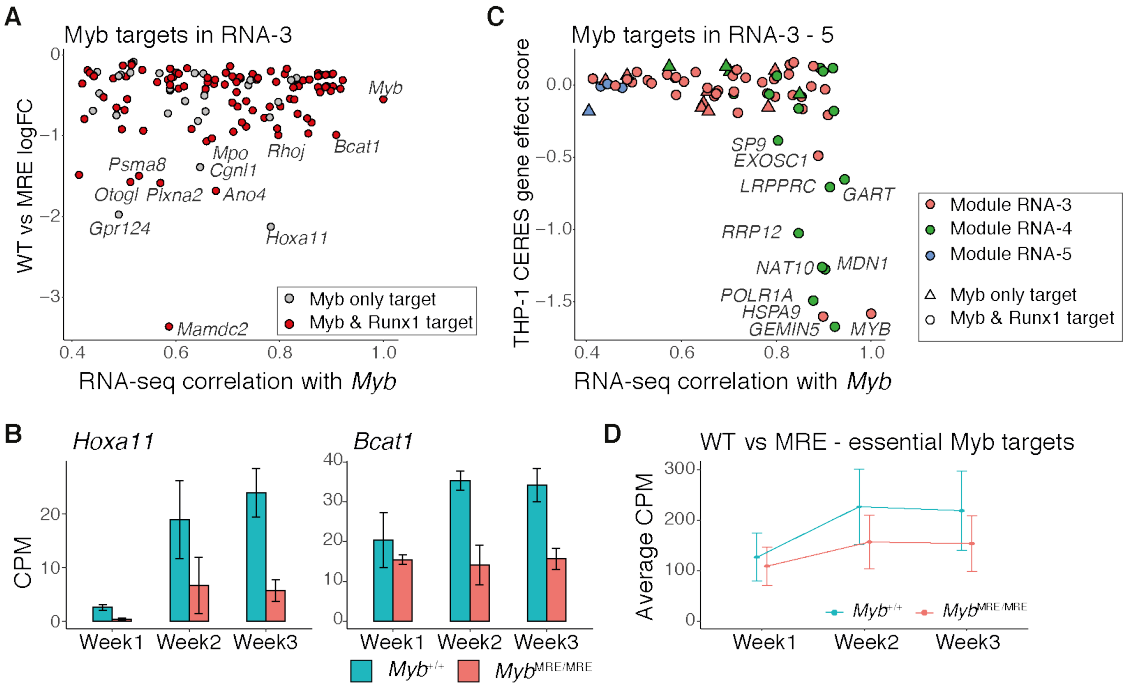


Figure 5.11: *Myb* is highly correlated with essential genes impacted by the $Myb^{MRE/MRE}$ circuit. (A) Association between RNA-seq logFC values comparing $Myb^{+/+}$ and $Myb^{MRE/MRE}$ and expression correlation with *Myb*. Data points shown are predicted Myb targets with FDR < 0.05 when comparing $Myb^{+/+}$ and $Myb^{MRE/MRE}$, and are coloured by whether genes are targeted by Myb only, or Runx1 and Myb. **(B)** Gene expression (CPM) for *Hoxa11* and *Bcat1*, as highlighted in A. Error bars represent standard error of the mean; $n = 3$. **(C)** Association between expression correlation with *Myb*, and THP-1 gene essentiality scores (CERES), where lower CERES scores indicate greater essentiality. Mouse genes were mapped to human gene symbols. Data points shown are predicted Myb targets in modules RNA3-5, with FDR < 0.05 when comparing $Myb^{+/+}$ and $Myb^{MRE/MRE}$. Genes are considered essential below CERES score -0.5. Colour represents gene expression module, while shape represents targeting by Myb only, or Runx1 and Myb. **(D)** Average gene expression (CPM) plots for all *Myb* correlated and essential genes highlighted in C. Data points represent mean CPM across all genes, and bars represent standard error of the mean.

While the $Myb^{MRE/MRE}$ genotype accompanies defects in leukaemic growth (Lau, 2022), it is as yet unclear which Myb interactions are critical for leukaemogenesis. To address this question, I estimated the importance of genes for leukaemogenesis with CRISPR essentiality data in THP-1 cells (MLL-AF9 AML cell line) from DepMap (22Q2 dataset, Doench *et al.*, 2016; Meyers *et al.*, 2017), using data for human orthologues of mouse gene targets. I compared CRISPR essentiality scores against *Myb* gene expression correlation at Myb target genes (Fig. 5.11C). This analysis was performed on genes downregulated in $Myb^{MRE/MRE}$ samples

in modules RNA3-5. Strikingly, the most essential Myb target genes were all highly correlated with *Myb* expression, while low or moderately correlated targets displayed no essentiality (Fig. 5.11C). Further, all of these essential genes were also predicted to be regulated by Runx1. This suggests that these essential genes are strongly associated with Myb expression levels, and may be supported by Runx1 activity. These genes are all expressed at a much lower level in $Myb^{\mathrm{MRE/MRE}}$ samples over $Myb^{+/+}$ (Fig. 5.11D). This analysis provides a mechanism by which *Myb* perturbation impairs leukaemogenesis, and highlights the association of Myb activity with essential downstream targets. It remains a question for future study as to whether Runx1 plays a role at these targets.

5.10 Conclusions and discussion

In this chapter, I have used a system established by I. Lau (Milne lab, Lau, 2022) to transfect mouse GMPs with an *MLL-AF9* expressing construct, and track gene expression changes over three weeks in culture. In doing so, I have established a set of experiments profiling gene expression, chromatin accessibility, and epigenetic modifications that monitor regulatory adaptations of *MLL-AF9* expression. By including $Myb^{\mathrm{MRE/MRE}}$ mouse models, these datasets also track how these cells adapt to *MLL-AF9* expression in the context of *Myb* perturbation. I have integrated this data together into an overall GRN model (section 5.4), that can be decoded to ask specific questions using gene expression modules (section 5.5). This data, while not yet fully explored, gives insight into how the wider regulatory network adapts to *MLL-AF9* expression, and how Myb activity affects leukaemogenesis.

In exploring RNA-seq and ATAC-seq data, it was found that the W1 time point was significantly different from W2 and W3, in both transcription and chromatin accessibility (section 5.3, Fig. 5.4). Interestingly, while W2 and W3 samples showed similar expression and accessibility profiles, it is at these time points that $Myb^{+/+}$ and $Myb^{\mathrm{MRE/MRE}}$ diverge. This suggested two points: 1). Gene expression and chromatin accessibility continues to adapt over time, long after *MLL-AF9* is first expressed, and 2). The conditional $Myb^{\mathrm{MRE/MRE}}$ perturbation does not affect regulation until W2. The transcription and accessibility changes from W1 to W2 were not mediated by a change in MLL-AF9 binding profiles, suggesting that this effect is independent of MLL-AF9, though H3K27 acetylation was reduced at W2 samples, indicating a change in the chromatin environment. It is important to note here that the ChIPmentation samples, while containing spike-in DNA, have not yet been reference normalised and therefore these results are not fully robust. Despite this caveat, these results suggest that gene regulation adapts to *MLL-AF9* expression in a gradual process despite the MLL-AF9 fusion protein being the sole driver of leukaemogenesis (Fig. 5.12).

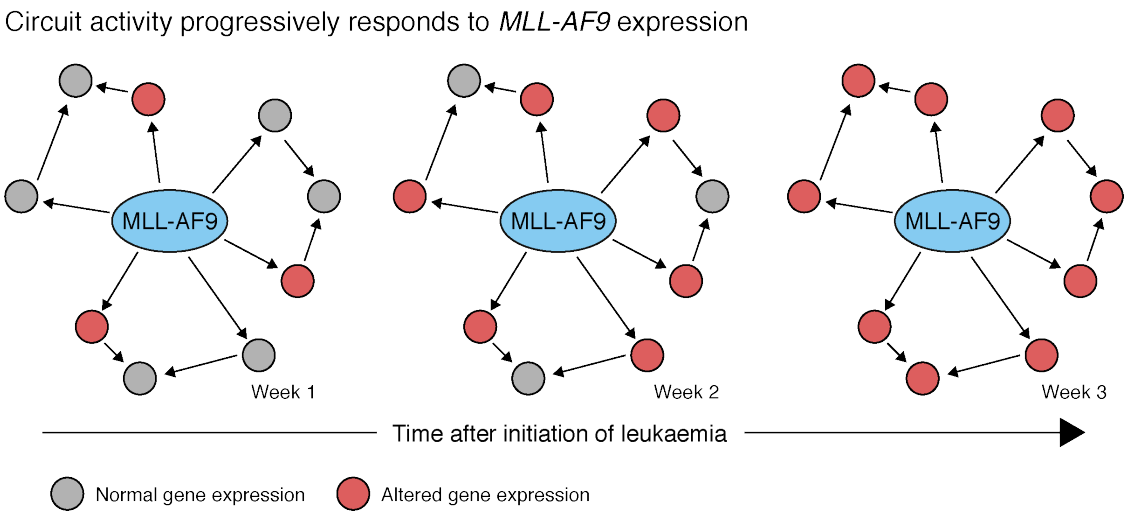


Figure 5.12: Summary model of how circuit activity progressively changes in response to MLL-AF9. A model illustrating how the TF network progressively responds to *MLL-AF9* expression over time. MLL-AF9 binding is stable at an early time point, and does not change further. Gene expression changes occur subsequent to MLL-AF9 binding, and continues to change over time. This suggests that transcriptional networks may gradually adapt their activity to the fusion protein, as opposed an immediate transformation of a normal GMP network.

The methodology used to construct the GRN model of MLL-AF9 GMP leukaemogenesis (section 5.4) was established using lessons learnt from the construction of the EHT GRN (chapter 3) and MLL-AF4 GRN (chapter 4). First, ChIPmentation data was used to identify H3K27 acetylation, and thus improves confidence that DAEs are enhancers (Creyghton *et al.*, 2010; Heintzman *et al.*, 2009; Heintzman *et al.*, 2007). Additionally, as opposed to relying on motif analysis, MLL-N ChIPmentation was used to evidence MLL-AF9 binding, though the ChIPmentation quality can be further improved with additional sequencing. The identification of E-P interactions in the EHT GRN used a basic approach by annotating accessible elements to the nearest differentially expressed promoter. In this chapter I built upon this method with an iterative approach that allowed for multiple E-P links for each enhancer. This method used an annotation window that expanded with each window until enhancer elements were sufficiently annotated, which allowed for genomic-context dependent annotation. Enhancers within gene dense regions would be sufficiently annotated in a smaller window than enhancers in gene sparse regions, preventing over annotation. As with the EHT GRN, upstream regulators were inferred by

motif analysis, and E-P links and predicted TF:target interactions were filtered on the basis of correlation. While the GRN model is not a perfect reflection of MLL-AF9 GMP biology, in particular due to the reliance on TF motifs, these steps improved the robustness of the GRN model for this chapter.

It is important to note that the MLL-AF9 GMP GRN model does not capture any data prior to the W1 time point. It is very likely that considerable effects would be seen by W1. However, the lack of transcriptional divergence between $Myb^{+/+}$ and $Myb^{\mathrm{MRE/MRE}}$ samples until W2 indicate that the time frame for this experiment captures a window of dynamic GRN adaptation. Generating additional RNA-seq and ATAC-seq data on baseline GMPs would improve the study, and allow comparisons with a normal GMP state. An additional consideration, is that the core consensus motif for Myb is small (5'-AACNG-3'), and as such frequently appears throughout the genome and may result in false-positive results. Therefore, the GRN model would benefit from additional Myb ChIPmentation data that would remove the reliance on motif analysis

The *Myc* gene is generally important across leukaemias, and as noted in chapter 4 (section 4.4.1, p.119) is essential in both MLL-AF4 and MLL-AF9 leukaemias. Further analyses in chapter 4 highlighted cooperation between RUNX1 and MYB to upregulate *Myc* expression (section 4.7, p.133). In the MLL-AF9 GMP GRN model *Myc* expression is not upregulated until W3 (section 5.6, Fig. 5.7), contrary to most genes which are upregulated at W2. As MLL-AF9 binding does not vary, this is likely due to the activation of TFs. Interestingly, while *Myb* expression is well correlated with *Myc*, *Runx1* is instead anti-correlated. The lack of strongly correlated upstream regulators suggests that the W3 *Myc* upregulation is driven by the activation of multiple TFs.

The observation that *Runx1* is anti-correlated with *Myc* expression highlights a striking difference between the MLL-AF4 ALL GRN and MLL-AF9 GMP GRN, in

that Runx1 activity appears to be different. While Runx1 activates *Myc* in MLL-AF4 ALL, it may repress *Myc* in MLL-AF9 GMPs. This has been observed in a study from Simeoni *et al.*, 2021, who propose that RUNX1 mediates *MYC* repression in AML by cooperating with FOXC3 to recruit a FOXC3/RUNX1/HDAC1/TLE3 repressor complex. This could be indicative of a general difference in Runx1 activity between myeloid and lymphoid lineages, or that this activity could be specific AML. It remains unclear as to whether Runx1 is a repressive TF genome wide, or if it shows context-dependent regulatory logic. As Myb frequently co-interacts with Runx1 in both the MLL-AF4 GRN (section 4.7, p.133) and the MLL-AF9 GMP GRN (section 5.7, Fig. 5.8), it is possible that Myb binding alters the regulatory logic of Runx1. This question is difficult to address without *Runx1* perturbation data, but it appears that genes predicted to be regulated by both Myb and Runx1 together are more sensitive to *Myb* perturbation ($Myb^{\mathrm{MRE/MRE}}$) than genes predicted to be regulated by Myb alone (section 5.9, Fig. 5.10). As such, it is possible that Runx1 supports the activity of Myb, though this hypothesis is as yet untested.

The GRN model was used to probe how Myb drives gene expression changes, and how Myb overexpression is important for leukaemic maintenance (Zuber *et al.*, 2011a) (section 5.9, Fig. 5.11). Using differential expression between the $Myb^{+/+}$ and $Myb^{\mathrm{MRE/MRE}}$ genotypes, in combination with *Myb* transcription correlation, *Hoxa11* and *Bcat1* were picked out as two example genes that are strongly downregulated in $Myb^{\mathrm{MRE/MRE}}$ samples, and are well-correlated with *Myb* expression. Myb driven upregulation of *Bcat1* could be an important interaction in MLL-AF9 leukaemias, as *Bcat1* overexpression is associated with poor AML prognosis, and is known to drive DNA hypermethylation similar to $\mathrm{IDH^{mut}}$ AML (Raffel *et al.*, 2017). Myb targets sensitive to *Myb* perturbation ($Myb^{\mathrm{MRE/MRE}}$) were also screened against DepMap gene essentiality data (essentiality in THP-1 cells, Doench *et al.*, 2016; Meyers *et al.*, 2017), and strikingly, all essential Myb targets were highly correlated with *Myb* expression ($R > 0.8$). This gene expression correlation suggests that Myb activity is tightly associated with the expression of these essential factors, and the statistical

comparison between *Myb*$^{+/+}$ and *Myb*$^{\mathrm{MRE/MRE}}$ genotypes suggests this relationship is functional. An essential target of particular relevance is *NAT10*, which is a known AML oncogene associated with poor prognosis, and upon perturbation leads to endoplasmic reticulum stress and apoptosis in AML (Liang *et al.*, 2020; Zi *et al.*, 2020). These analyses position Myb as an important TF that drives regulation of multiple critical genes in the MLL-AF9 GRN.

6

General discussion

Contents

6.1 TF cooperation underpinning normal and aberrant cell processes 180
6.2 The deregulation of normal GRNs in *MLL*r leukaemia. 182
6.3 Future work 185
6.3.1 Validating predicted interactions 185
6.3.2 Investigating TF cooperativity 186
6.3.3 Improving GRN analyses 186

The specific aims of this study were to identify novel regulatory mechanisms driving key biological processes, to explore TF cooperation using a GRN framework, and to probe GRN models of *MLL*r leukaemia and determine how normal circuits are deregulated. This study was applied to mouse EHT as an example of a normal haematopoietic process, and MLL-AF4 and MLL-AF9 leukaemias as genetically simple transcriptional diseases. In doing so I applied different methodologies for creating a GRN model, as well as using different approaches for analysing the resulting networks. In particular, I tested an approach for investigating the frequency with which TFs co-interact at the same loci, which successfully predicted both known and novel cooperative TF pairs. This investigation revealed insights into the GRNs underpinning EHT and *MLL*r leukaemias as driven by TF cooperation, highlighting the utility of GRN models for probing the combinatorial logic of TFs, as well as elucidating general principles by which regulatory circuits are disrupted in disease.

6.1 TF cooperation underpinning normal and aberrant cell processes

In this study I have used GRN models to investigate the regulatory interactions underpinning EHT and *MLL*r processes, using a strategy to identify probable TF co-interactions (methods section 2.15.6, p.60).

This strategy identified multiple TFs predicted to co-interact with Runx1, with the potential to set up FFLs. These included Ikzf1, Ets1, Meis1, Nfe2, PU.1, Mef2c, Gfi1, Pbx1, and Cebpb. These TFs were associated with motifs enriched for different modules of the GRN, such as a Meis1 and Nfe2 motif bias for endothelial enhancers (ATAC2), which might suggest that Runx1:Meis1 and Runx1:Nfe2 FFLs regulate an endothelial program. A number of these co-interacting factors were associated with motifs enriched in transiently accessible enhancers, such as Mef2c, Gfi1, and Ikzf1, suggesting these factors regulate the transient network. Taking the Notch pathway as a key transient process, I identified a Runx1:Ikzf1:*Hes1* repressive FFL, where

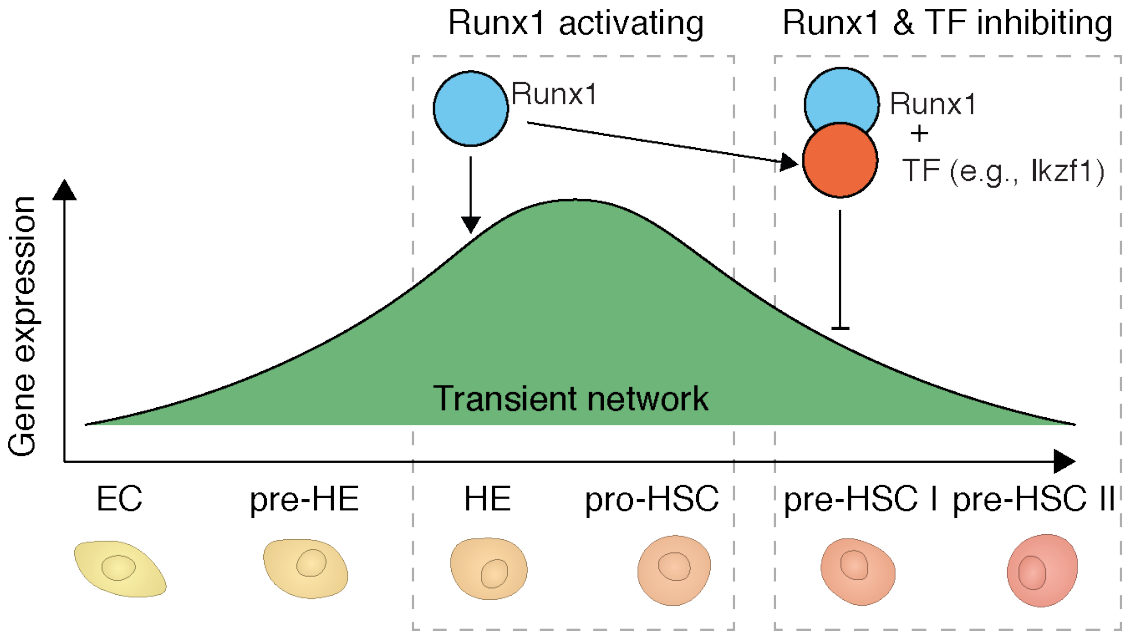


Figure 6.1: Hypothetical model describing Runx1 regulation of the transient EHT network. A model describing the hypothesis of Runx1 regulation of the transient EHT network. *Runx1* initially correlates with the expression of transiently expressed genes, and Runx1 may activate the transient network. The transient genes are downregulated with the onset of late-EHT genes, such as *Ikzf1*. As preliminary evidence suggests Runx1 activity swaps to repression when *Ikzf1* is expressed (Appendix A.14), this could indicate a mechanism by which late-EHT TFs alter Runx1 activity to then deactivate the transient network.

Runx1 and Ikzf1 are predicted to act cooperatively (section 3.6.1, p.90). Existing literature suggests a physical interaction between Runx1 and Ikzf1 (Zhou *et al.*, 2019), and data from L. Greder (de Bruijn lab, unpublished) indicate that Runx1 swaps from an activating TF, to repressive when *Ikzf1* is expressed (Appendix A.14). This could be extrapolated as a general mechanism in which Runx1 behaviour is altered. It is possible that in early-EHT Runx1 is an activator, correlating with the onset of the transient EHT network (section 3.2.1, p.66). Runx1 could promote the transient expression of genes, such as *Hes1*, and as Runx1 drives the expression of late-EHT genes (i.e., *Ikzf1*), Runx1 activity may switch to repressive, subsequently inhibiting the transient network (Fig. 6.1). This cooperative behaviour may also exist for other Runx1 co-interacting TFs, such as Mef2c, which is also associated with a motif highly enriched in transiently accessible enhancers. While this is based on observational data and has not been extensively tested, it is an interesting hypothesis

that could suggest a previously unappreciated role in Runx1 control of EHT.

In *MLL*r leukaemia, TF cooperation was investigated by exploring how RUNX1 and MYB co-regulate targets in the context of MLL-AF4 binding. RUNX1 and MYB were found to function together to increase expression of *BCL2* and *MYC* (section 4.7, p.133). MLL-AF4 is the primary driver of *MYC* and *BCL2* expression, but the cooperative action of RUNX1 and MYB incrementally increased expression of these loci. As it also appears that MLL-AF4 encourages the binding of TFs (section 4.6.2, p.130), it indicates that MLL-AF4 functions cooperatively with multiple TFs, and highlights the importance of incremental TF activity underpinning MLL-AF4 target genes.

6.2 The deregulation of normal GRNs in *MLL*r leukaemia.

One of the main questions asked in this thesis is "how are normal circuits altered in disease?" This was largely addressed through analysis of MLL-AF4 ALL and MLL-AF9 AML GRN models in chapters 4 and 5. The analysis in these chapters hint towards several possible mechanisms by which MLL-FPs could transform a normal function network. For instance, MLL-AF4 binding and spreading increases the frequency of RUNX1 and MAZ binding (section 4.6.2, p.130). This appears to be mediated through increased affinity for existing TF DNA binding, which may be driven through MLL-AF4 altering the chromatin environment. This could be a general principal for all TFs, whereby MLL-AF4 co-opts the activity of existing TFs and directs their activity to MLL-AF4 targets. As such, this provides a novel model for how MLL-AF4 sets up novel disease-specific interactions by altering the chromatin environment (Fig. 6.2).

One interesting observation is that RUNX1 activity appears to change depending on the presence of MLL-AF4 (section 4.6, p.126). It was observed that at gene targets with RUNX1 but no MLL-AF4, RUNX1 both upregulated and downregulated

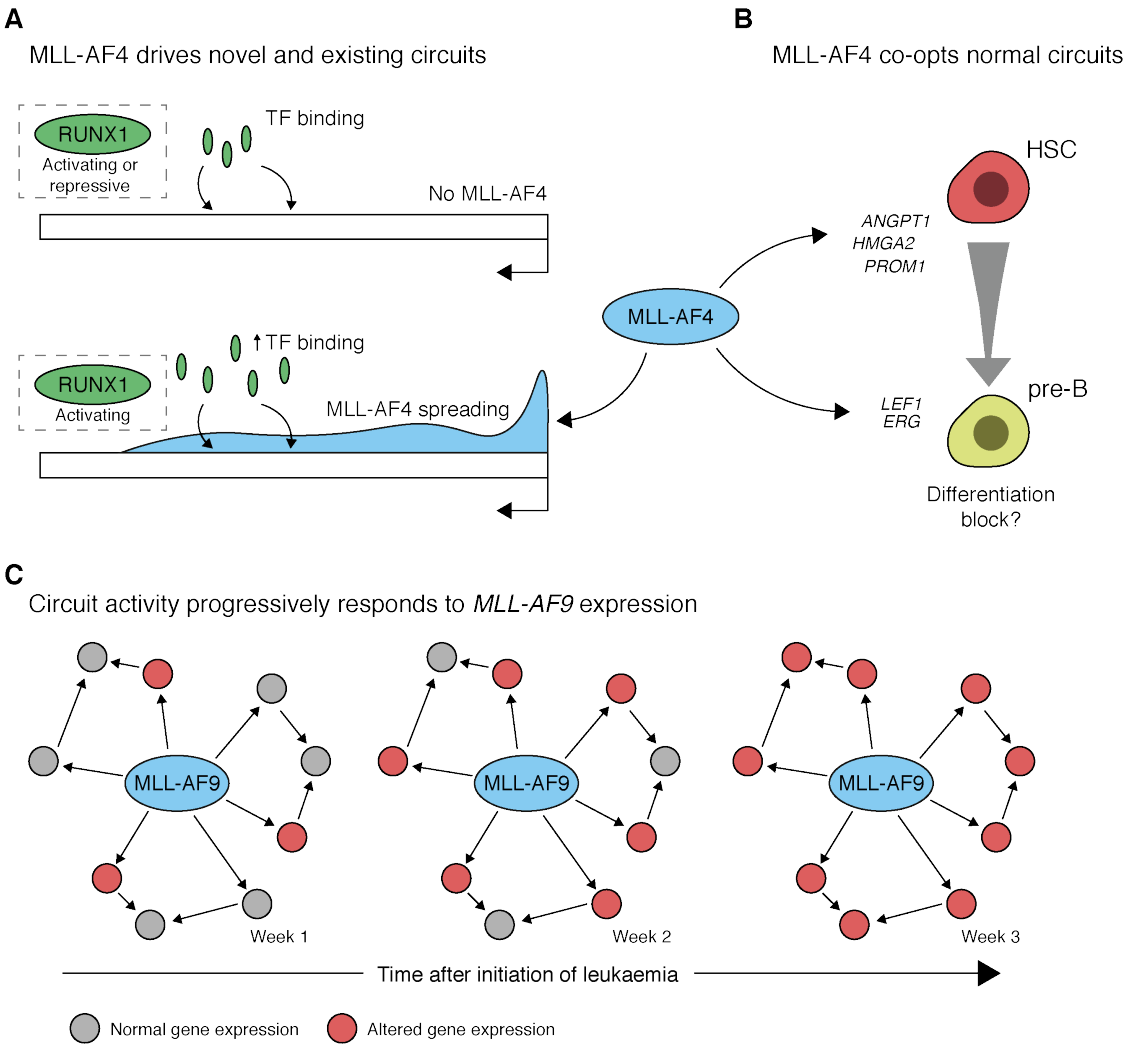


Figure 6.2: Summary model describing possible mechanisms by which the normal GRN is affected in *MLL*r leukaemia. The impact of MLL-AF4 or MLL-AF9 on a normal transcriptional network can be described as three general concepts. **(A)** First, MLL-AF4 has the potential to establish novel TF interactions and alter the activity of RUNX1. **(B)** Secondly, MLL-AF4 can co-opt pre-existing normal circuits, either to reactivate early HSC expressing genes, or to forcibly maintain expression of pro-B or pre-B genes. This co-opting of circuits could be underlie the differentiation block of *MLL*r ALLs. **(C)** Additionally, it appears that the activity of circuits do not immediately change upon expression of *MLL-AF9*, and that instead gene expression is altered over time.

expression. This is congruent with the known behaviour of RUNX1, which can recruit both activating and repressing cofactors (Canon *et al.*, 2003; Kitabayashi *et al.*, 1998; Perry *et al.*, 2002; Yamagata *et al.*, 2005). However, where MLL-AF4 is bound, RUNX1 activity is biased towards activation. This suggests that MLL-AF4 alters the activity of RUNX1, potentially in a similar fashion to the Runx1 and

Ikzf1 cooperation as discussed in section 6.1. This highlights the potential for MLL-AF4 to not only promote novel GRN interactions, but to alter the behaviour of TFs to further activate genes (Fig. 6.2).

While MLL-AF4 appears to alter the normal GRN through novel interactions and biasing TF regulatory logic, it is clear that it also co-opts and activates existing circuits. This was investigated by comparing the MLL-AF4 GRN model with fetal B lymphopoiesis signatures (section 4.8, p.136), where MLL-AF4 was predicted to interact and drive several B lymphopoiesis associated genes. *ERG* is normally expressed from HSCs through to pro-B cells, and is strongly upregulated by MLL-AF4. I speculated that MLL-AF4 overexpression of *ERG* expression could explain the block in differentiation seen in *MLL*r ALL (Lin *et al.*, 2016), as *ERG* overexpression causes a partial pro-B to pre-B block in mice (Tsuzuki *et al.*, 2011) (Fig. 6.2). Additionally, this analysis noted activation of stem cell genes, such as *HMGA2* and *ANGPT1*, both of which have significant regulatory roles and contribute to leukaemogenesis (Eguchi-Ishimae *et al.*, 2014; Guenther *et al.*, 2008; Zhou *et al.*, 2015). This highlights the capacity for MLL-AF4 to reactivate stem cell regulatory factors not normally present in pro-B cells (Fig. 6.2). It is interesting that in this analysis only a few stem cell genes are predicted targets of MLL-AF4, suggesting the fusion protein can only reactivate (or maintain expression, depending on the cell type the translocation event occurs) a specific subset of genes. What discriminates the targeting of these genes from others is not yet clear.

A final observation was made using the MLL-AF9 GRN, which incorporated a time course experiment following leukaemic transformation. It was noted that despite a lack of change in MLL-AF9 binding over the time course, there were significant gene expression and chromatin accessibility changes (section 5.3, p.157). This suggested that while MLL-AF9 likely binds to gene targets rapidly, the impact on gene expression is progressive. This was most striking with the observation that the transcriptome of $Myb^{\mathrm{MRE/MRE}}$ mice did not diverge from $Myb^{+/+}$ mice until two weeks after leukaemia initiation. This experiment does not contain baseline

levels of expression in untransformed GMPs, so the most immediate transcriptional effects are lost, however it is clear that these GRNs adapt in complex ways over time, likely due to secondary effects resulting from initial *MLL-AF9* expression.

6.3 Future work

6.3.1 Validating predicted interactions

In chapter 3 a number of candidate drivers of EHT were predicted, including Ikzf1, Cd109, Maf and Lef1, among others. As these are predictions based on the GRN model, functional validation is necessary to draw conclusions on EHT progression. A straightforward approach to do so is to knockout these factors in mESCs, followed by haematopoietic differentiation to assay the impact on EHT. However, mESC differentiation is not a perfect model of in vivo EHT, and therefore to truly validate these drivers mouse knockout models are required. For predicted regulators of the EHT budding process, such as Maf and Lef1, it would be interesting to observe the effect of their perturbation on cytoskeletal structures.

For chapters 4 and 5 the GRN models predicted several direct and indirect targets of MLL-AF4 and MLL-AF9, that are proposed to be important for leukaemia function. These factors could simply drive survival and proliferation, but could also contribute to several additional aspects of cancer biology. For instance, while a knockout of a key TF may not impact cell growth in culture, it may affect bone marrow engraftment in transplant experiments, or alter the bone marrow niche environment as seen with *ANGPT1* (Zhou *et al.*, 2015). Therefore, to truly validate the importance of key nodes in the *MLL*r GRNs mouse transplants would be ideal. An interesting model of infant MLL-AF4 leukaemia has been established in Rice *et al.*, 2021, where an *MLL-AF4* translocation can be induced CD34+ FL cells via CRISPR Cas-9, then transplanted into mice. Using this model to assay specific TF perturbations would be a convincing validation of the importance of these key nodes.

6.3.2 Investigating TF cooperativity

Several cases of TF cooperativity have been proposed, such as where Runx1 cooperation with Ikzf1 and MLL-AF4 is implied to change Runx1 activity, and warrants further investigation. A system developed by Blackledge *et al.*, 2014 can be used to study protein interactions in mESCs, referred to as the Tet repressor (TetR) system. TetR mESCs have an inserted Tet operator (TetO) array embedded in a 170 kb gene desert. By expressing *TetR* fused to a gene of interest, the resulting TetR:protein fusion binds to the transcriptionally inactive TetO, allowing us to probe the recruitment of cofactors. By generating TetR:Runx1 fusions, alongside expression of *Ikzf1*, the recruitment of cofactors can be investigating in the context of Ikzf1 activity. Alternatively, the perturbation studies can be redesigned. Knockout and siRNA knockdown experiments involve perturbation treatments lasting at least multiple days, or even incomplete removal of protein. The dTAG system involves fusing a $FKBP^{F36V}$ to the N- or C-terminus of a protein of interest, then upon treatment with small molecule dTAG-13, the protein-$FKBP^{F36V}$ fusion is targeted by the CRBN E3 ligase complex and rapidly degraded (Nabet *et al.*, 2018). This can be used to rapidly degrade Runx1, Ikzf1, or MLL-AF4, and allow us to determine the impact of specific circuits on gene regulation in a much shorter time frame, thus presenting cleaner evidence of TF cooperation dynamics in FFLs with fewer secondary effects following perturbation.

6.3.3 Improving GRN analyses

The analysis in chapter 5 is largely preliminary, and additional work is required to fully investigate this system. For instance, an important caveat for this chapter is that the Myb TF has a short core motif (5'-AACNG-3', Ogata *et al.*, 1994), which results in an abundance of matching DNA sequences across the genome. The data could therefore benefit from Myb ChIPmentation for stronger evidence of Myb binding. Additionally, the analysis has strongly focused on just a few gene expression modules that highlight overexpression in $Myb^{+/+}$ over $Myb^{MRE/MRE}$

genotypes, and an exploration of the wider GRN could reveal interesting mechanisms of GRN evolution over time.

In general, there are a number of additional GRN methodologies that could be applied to the systems studied in this thesis. More advanced alternatives to measures of centrality have been developed that attempt to measure the influence of a node (node influence metric). One example is the measure of "expected force" (Dunn *et al.*, 2005; Girvan *et al.*, 2002), which quantifies the information spreading potential of each GRN node. Additionally, more advanced network-based methods of identifying specific GRN modules could be applied, as opposed to clustering based on gene expression patterns. For example, the edge-betweenness approach applies the betweenness centrality calculation to edges rather than nodes, then high betweenness edges are iteratively deleted, until distinct, disconnected networks are established. This is an alternative method to identify complex structures in the GRN models, and could highlight circuits underpinning key processes that have otherwise been missed using a gene expression clustering approach.

Gene regulation studies have increasingly moved towards single-cell based approaches, with recent technological advancements allowing both RNA and ATAC-seq simultaneous profiling. The use of single cell approaches for development of networks is an exciting prospect, as specific circuits may exist in particular cell types, or expression of TFs may be stochastic or oscillate, which single cell resolution assays may offer insight into. The RNA+ATAC multiome data used in section 4.8 p.136 could be used to establish both ALL and normal B-lymphopoiesis GRN models. Tools have been developed to create GRN models from these data, such as the recently updated SCENIC+ package (González-Blas *et al.*, 2022). Additionally, Perturb-seq is a single cell approach that has been developed to integrate a genome wide CRISPR screen paired with gene expression profiling (Dixit *et al.*, 2016), thus allowing the impact of many gene perturbations on the surrounding network. This has been applied recently to the K562 cell line (chronic myeloid leukaemia) (Replogle *et al.*, 2022), establishing a genome wide genotype-phenotype map, that

could be a strong resource for establishing GRN models. The use of these single cell methods is an exciting prospect, and could offer insights into the dynamics of GRN circuits at a single cell resolution.

Appendices

A

Appendix

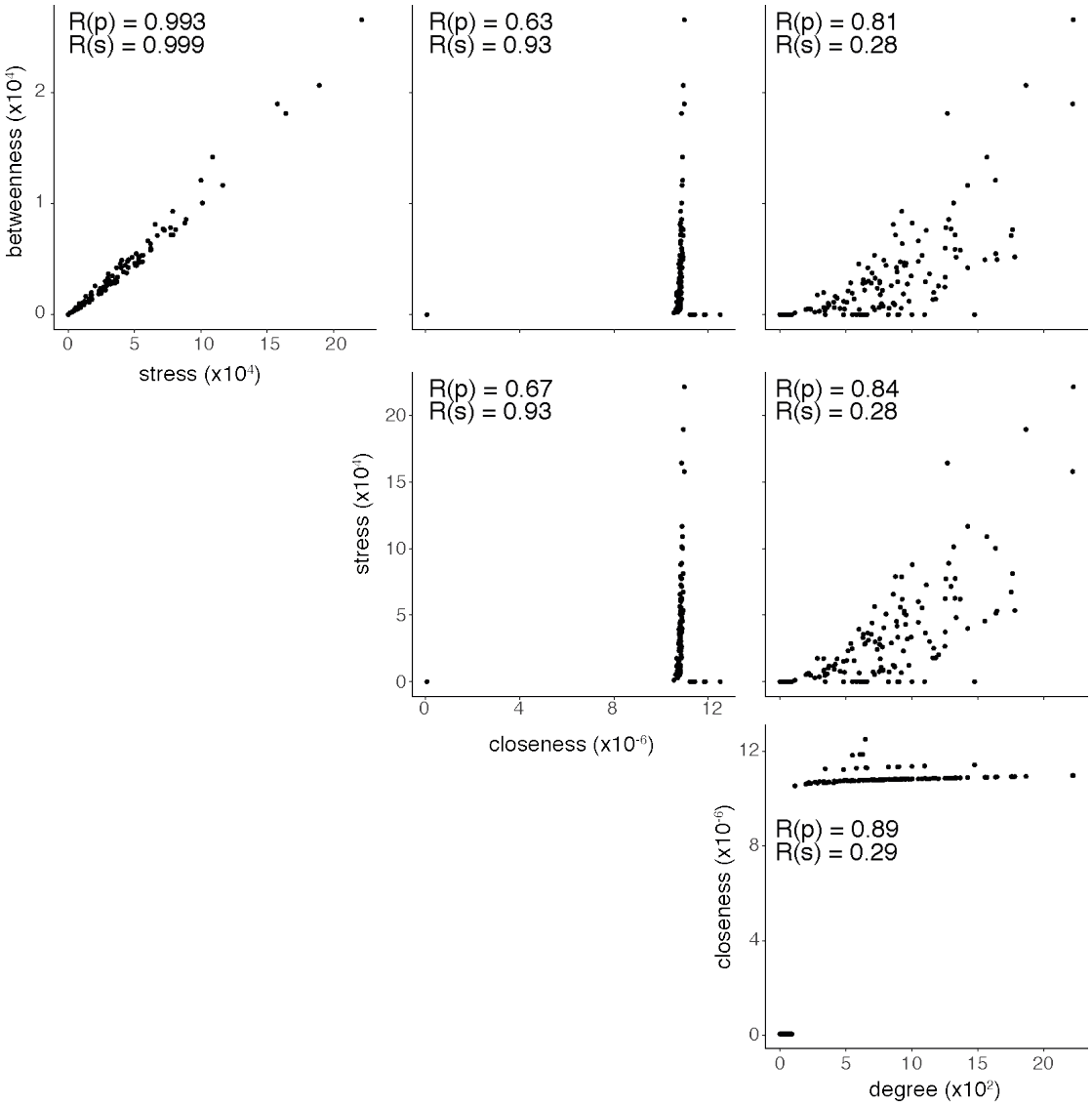


Figure A.1: Relationship between different centrality measures. Relationship between different centrality measures, as calculated using the EHT GRN.

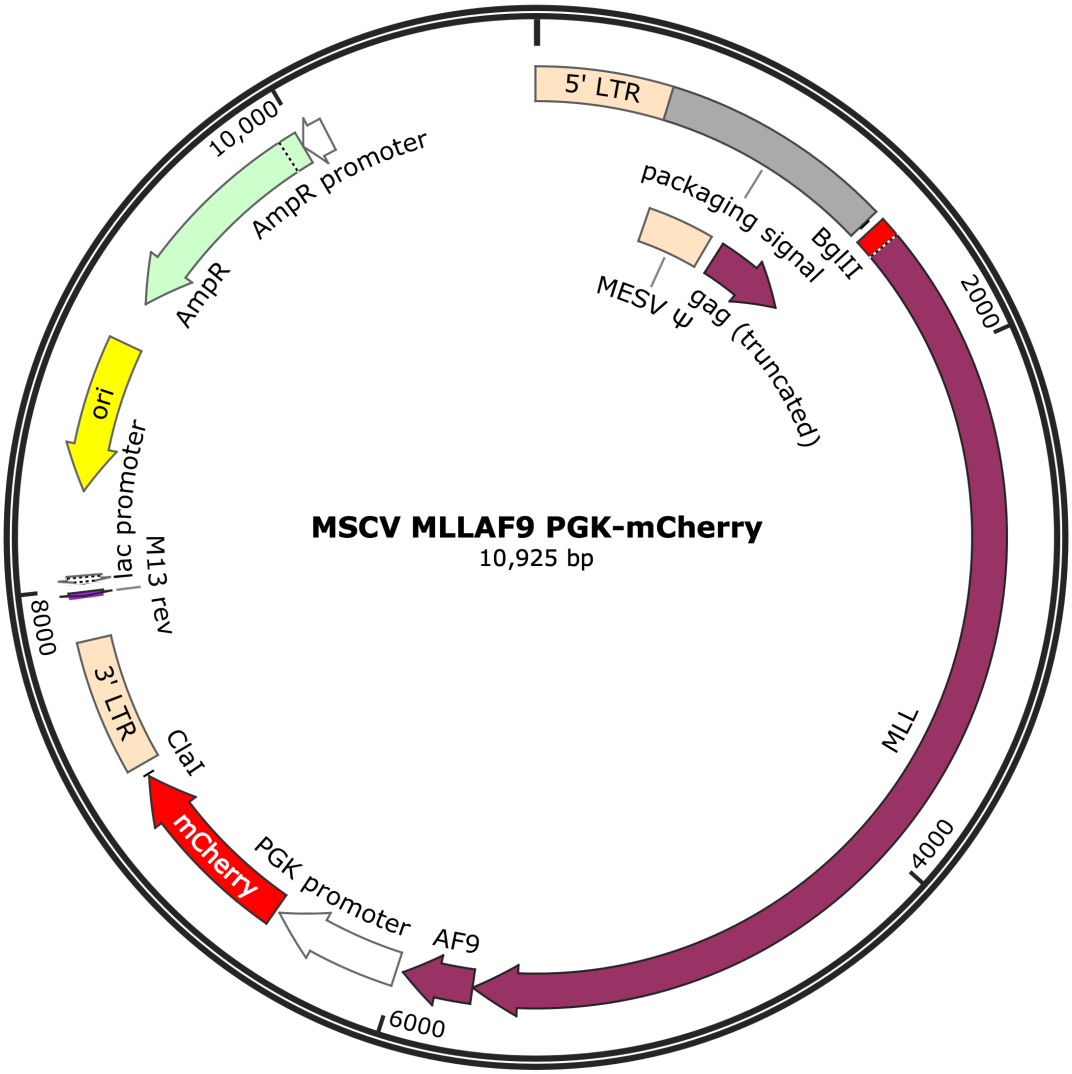


Figure A.2: Vector map of MSCV MLL-AF9 PGK-mCherry. *Vector and vector map generated by I-Jun Lau.*

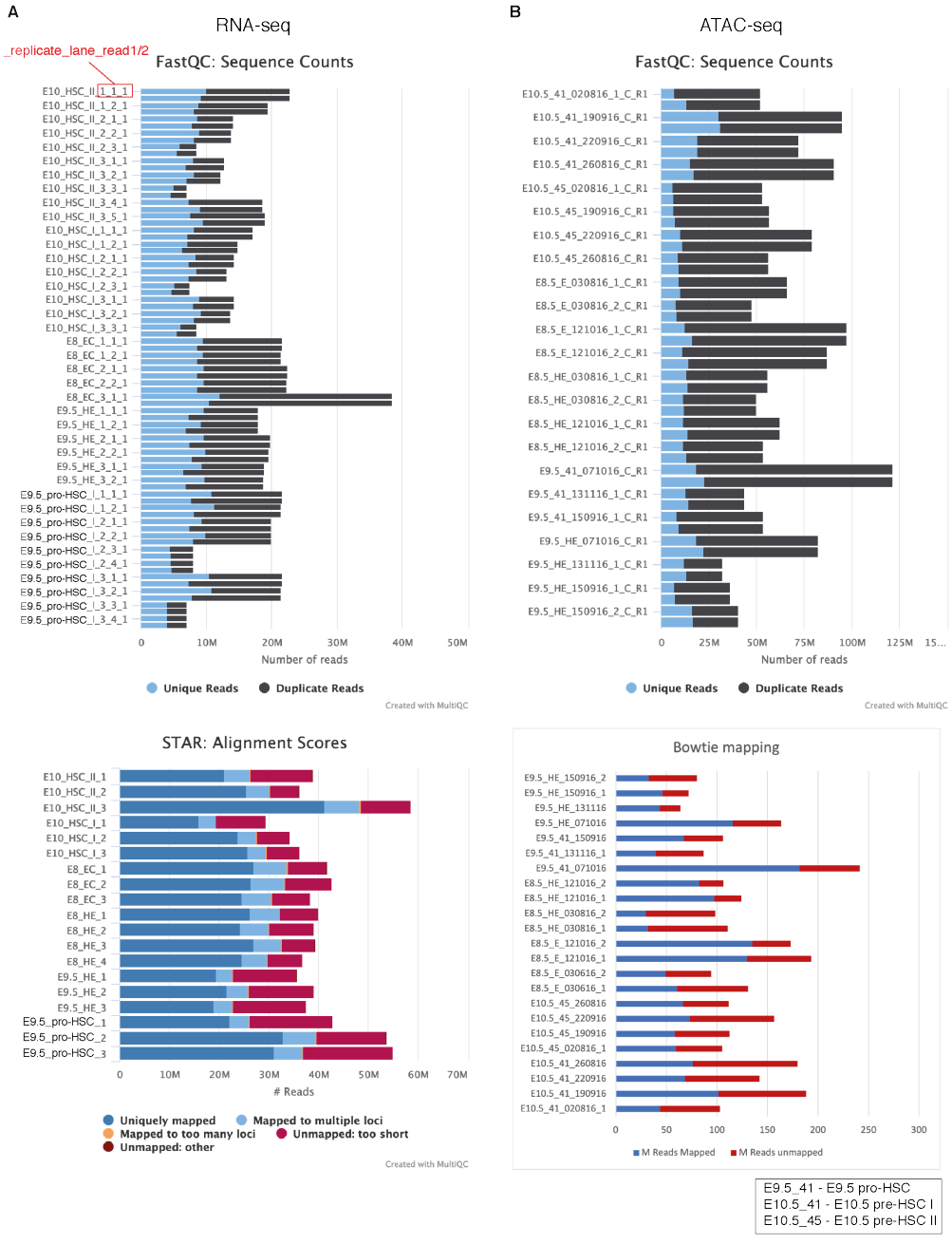


Figure A.3: Quality control plots for EHT RNA-seq and ATAC-seq. (A-B) FastQC breakdown of read duplication, STAR mapping metrics, and Bowtie2 mapping metrics for RNA-seq (A) and ATAC-seq (B).

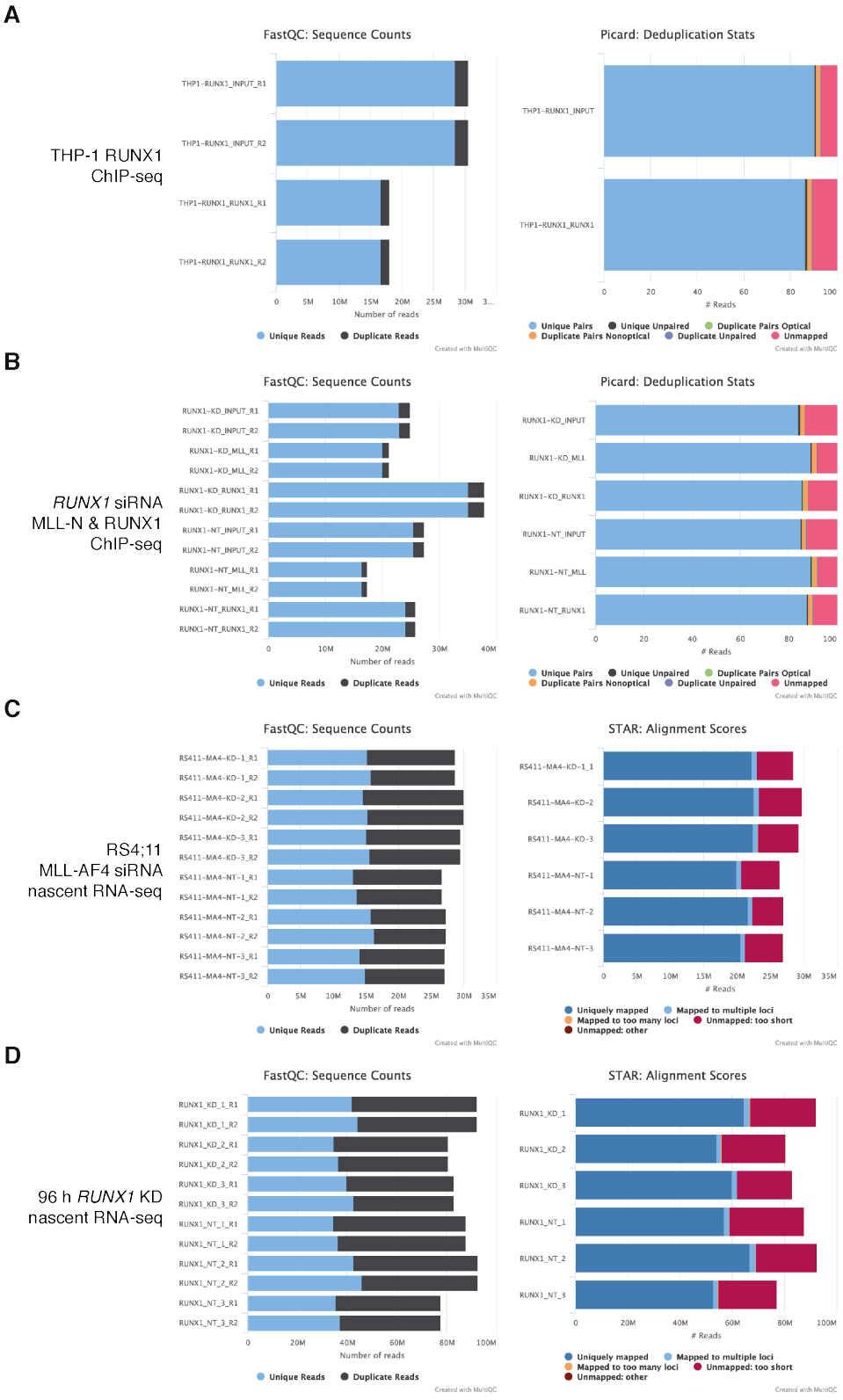


Figure A.4: FastQC and mapping statistics for unpublished datasets in chapter 4. *Legend continued on next page.*

Figure A.4: *Legend continued from previous page.* FastQC (left) and read mapping statistics (right) for THP-1 RUNX1 ChIP-seq (A), SEM 96 h *RUNX1* KD MLL-N and RUNX1 ChIP-seq (B), RS4;11 96 h *MLL-AF4* KD nascent RNA-seq (C), and SEM 96 h *RUNX1* KD nascent RNA-seq (D). *Nascent RNA-seq data in (D) was generated by M. Tapia (Milne lab), and analysed by me.*

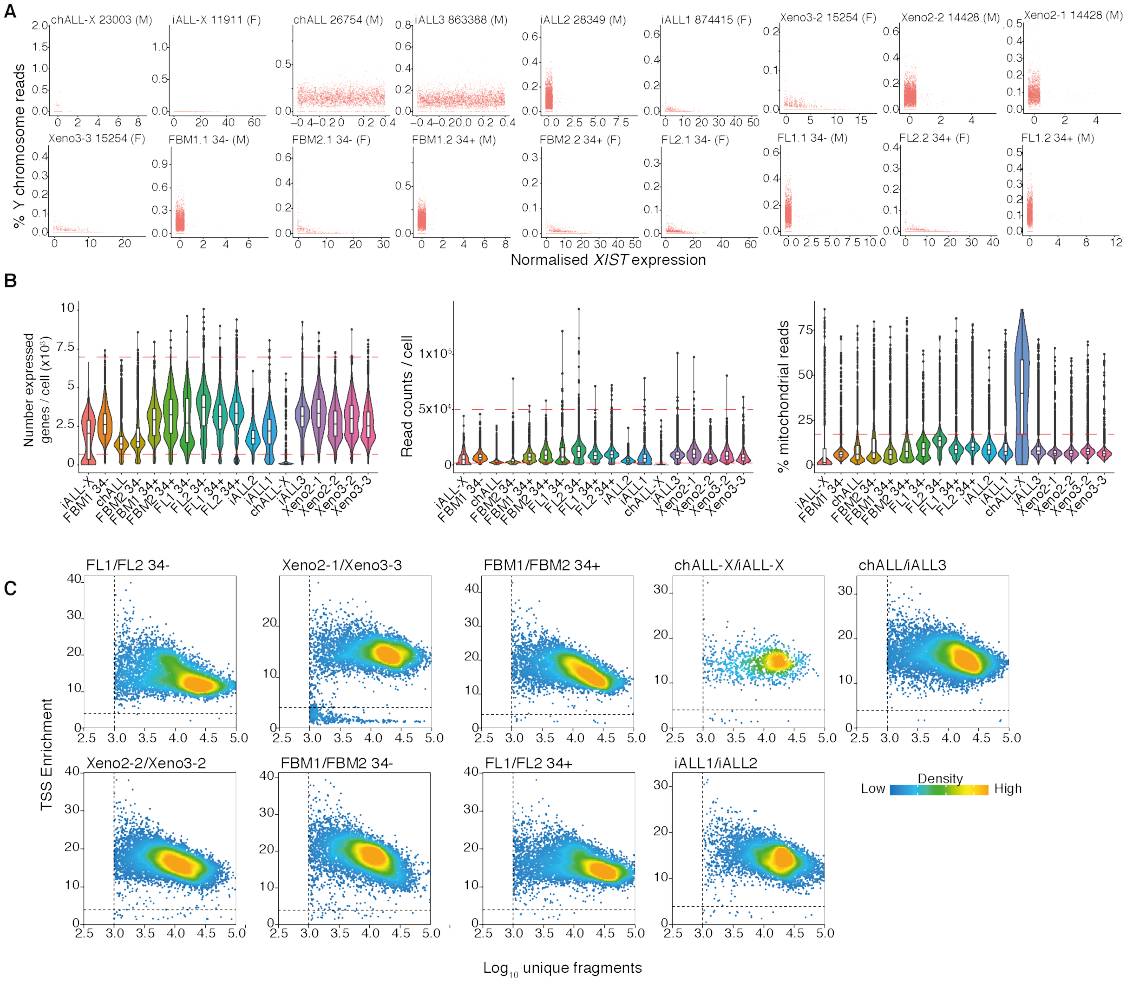


Figure A.5: RNA+ATAC multiome quality control plots. (A) Single cell Feature plots showing *XIST* expression against percentage of reads in Y chromosome. Data used to judge gender of the population, for deconvoluting multiplexed samples. **(B)** Violin and boxplots showing distribution of total detected genes, total read count, and percentage mitochondrial reads. Red dashed lines indicate quality thresholds (see methods). **(C)** ArchR quality control plots, showing total unique fragments against TSS enrichment. Dashed line indicates quality thresholds (see methods). *Plots and data filtering in C performed by Alastair Smith.*

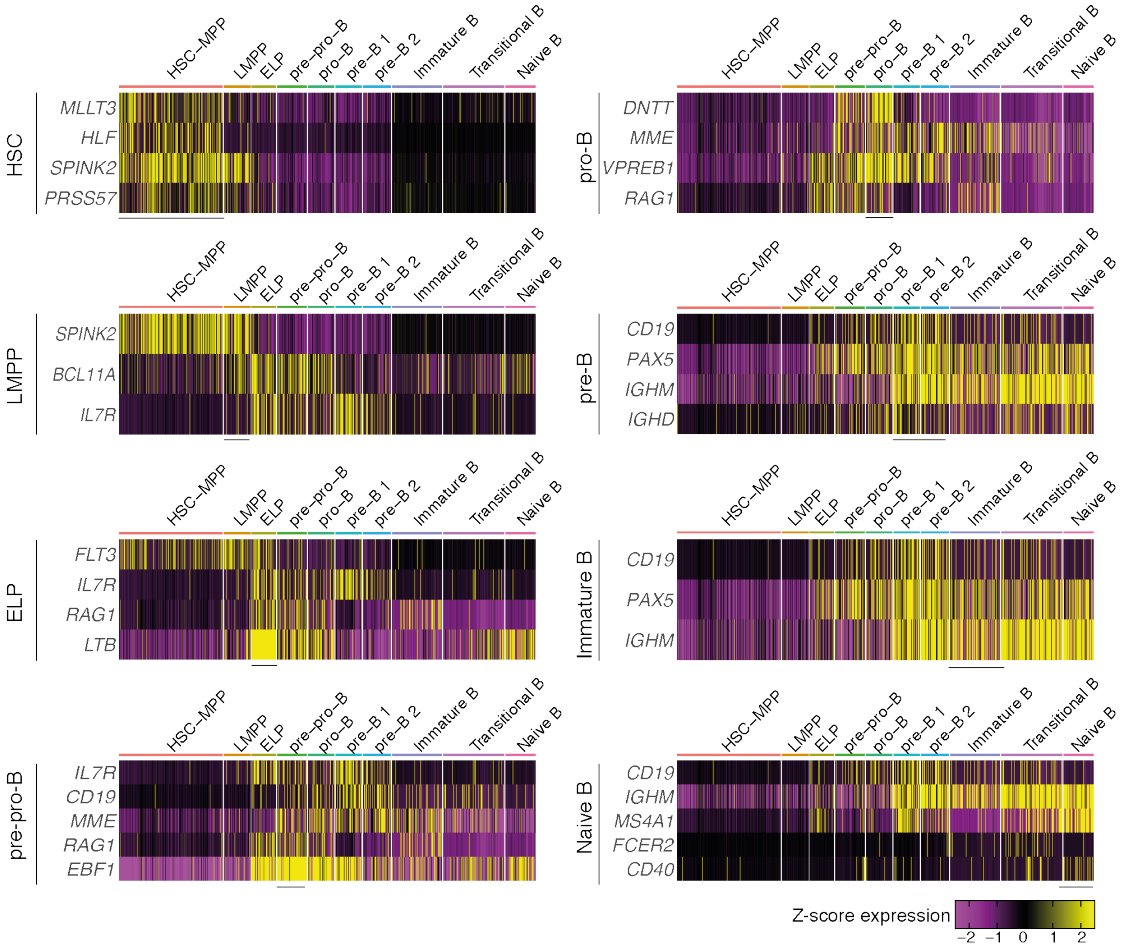


Figure A.6: Sc-multiome lineage markers defining B progenitor populations. Heatmap of z-score scaled normalised gene expression over labelled B progenitor populations. Genes shown are transcriptional markers differentiating B progenitor populations (Jardine *et al.*, 2021; Suo *et al.*, 2022). Note transitional B population is defined as such as it is a cluster between immature and Naive B populations.

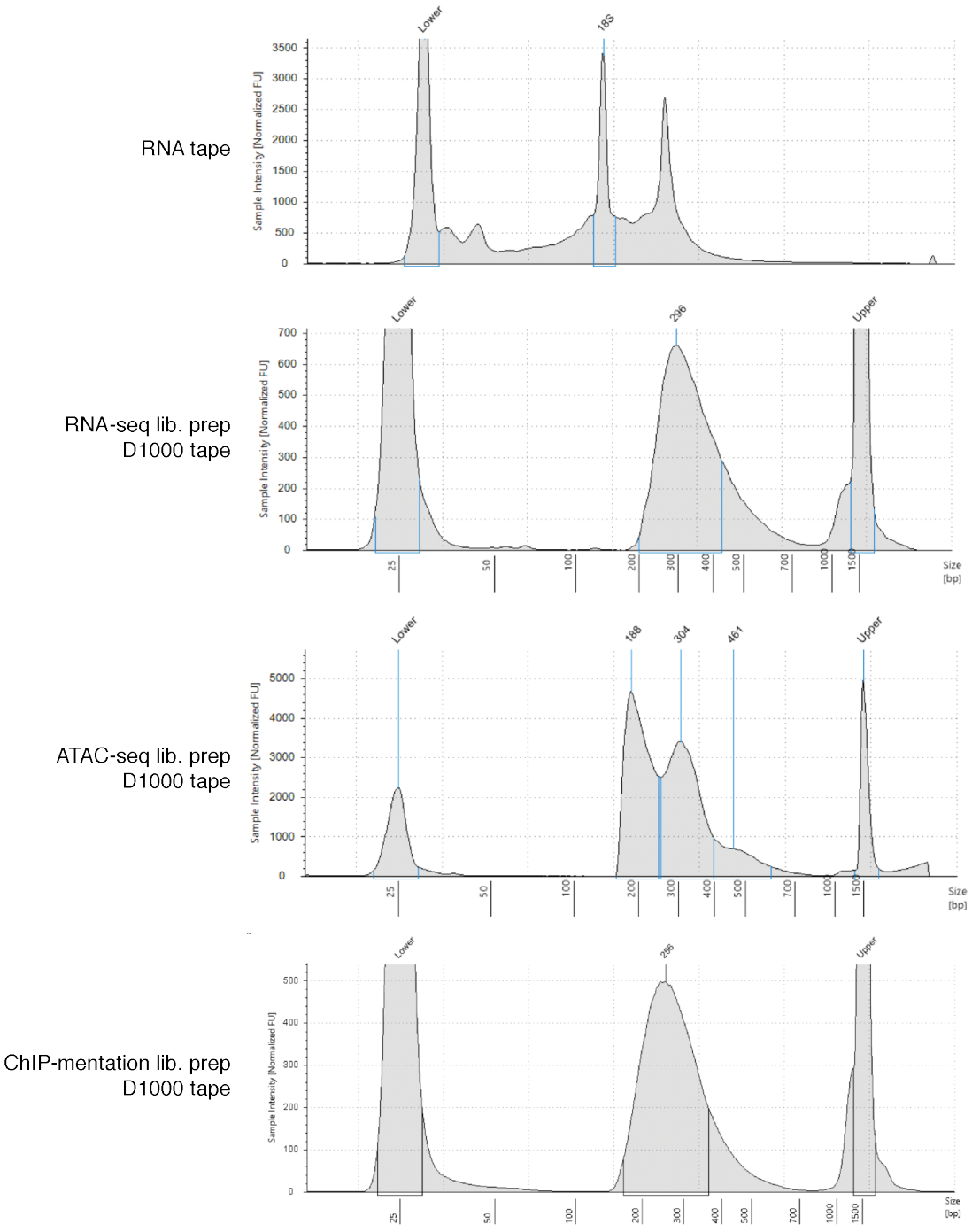


Figure A.7: Tapestation traces for MLL-AF9 GMP MethoCult replating. Representative tapestation traces for RNA (top), and RNA-seq, ATAC-seq and ChIP-mentation libraries (bottom), generated from the MLL-AF9 GMP MethoCult replating experiment.

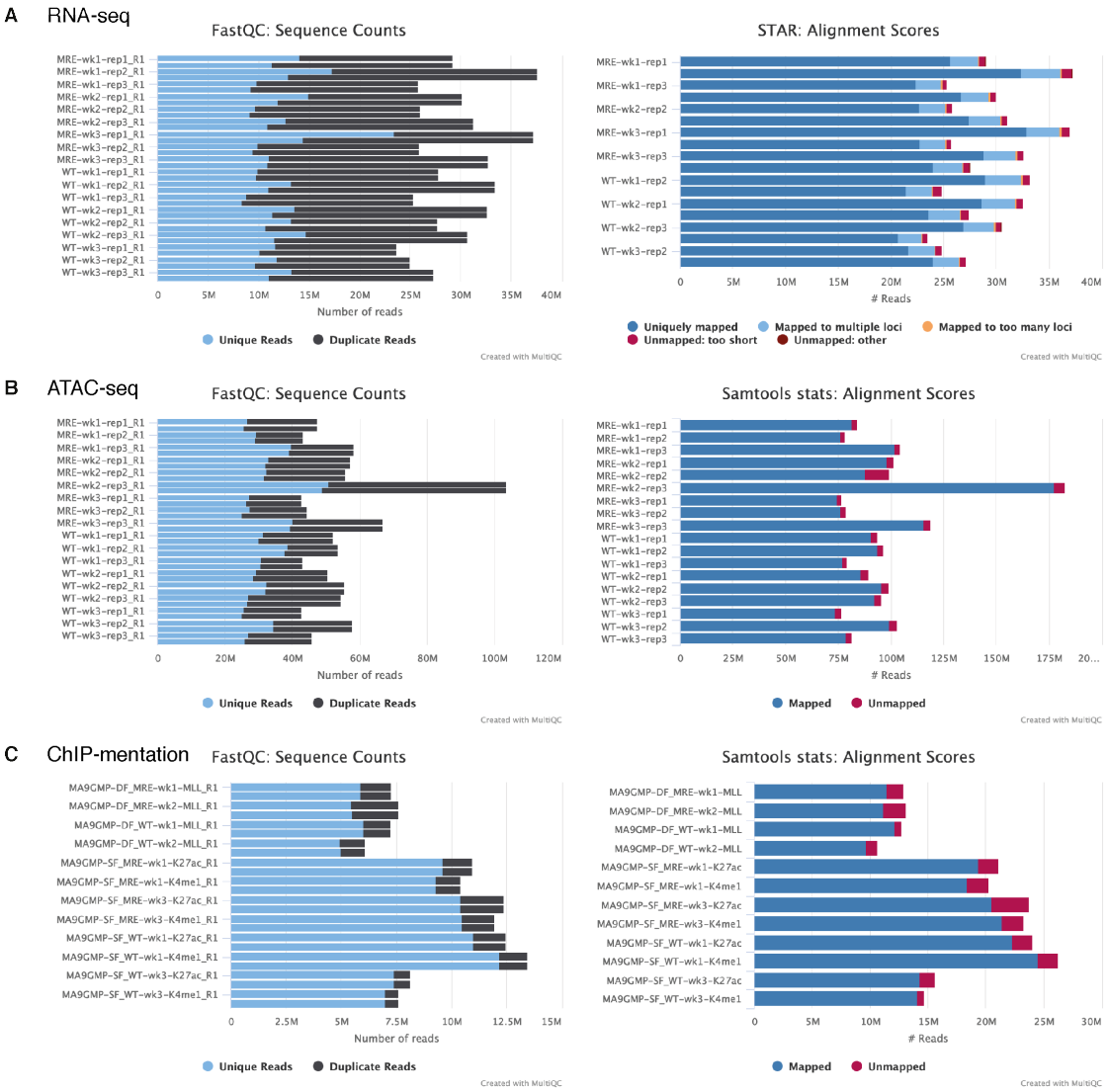


Figure A.8: FastQC and mapping statistics for MLL-AF9 GMP MethoCult replating. FastQC (left) and read mapping statistics (right) for RNA-seq (A), ATAC-seq (B) and ChIP-mentation (C) data, generated from the MLL-AF9 GMP MethoCult replating experiment.

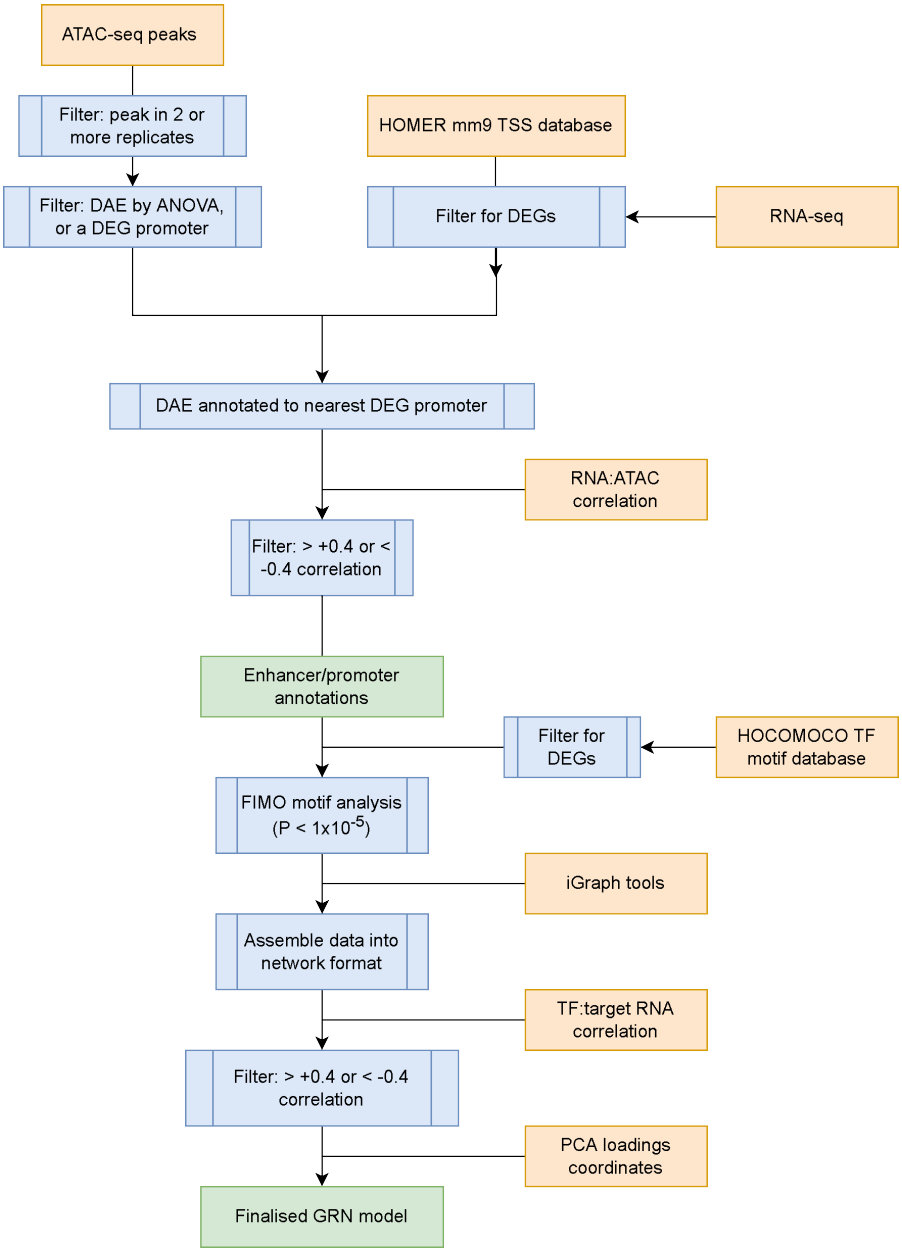


Figure A.9: Flow diagram explaining the integration of RNA-seq and ATAC-seq data to generate the EHT GRN model. Flow diagram schematic outlining the steps involved in the integration of RNA-seq and ATAC-seq data to generate the EHT GRN. Input data coloured orange; data processing steps coloured blue; data outputs coloured green.

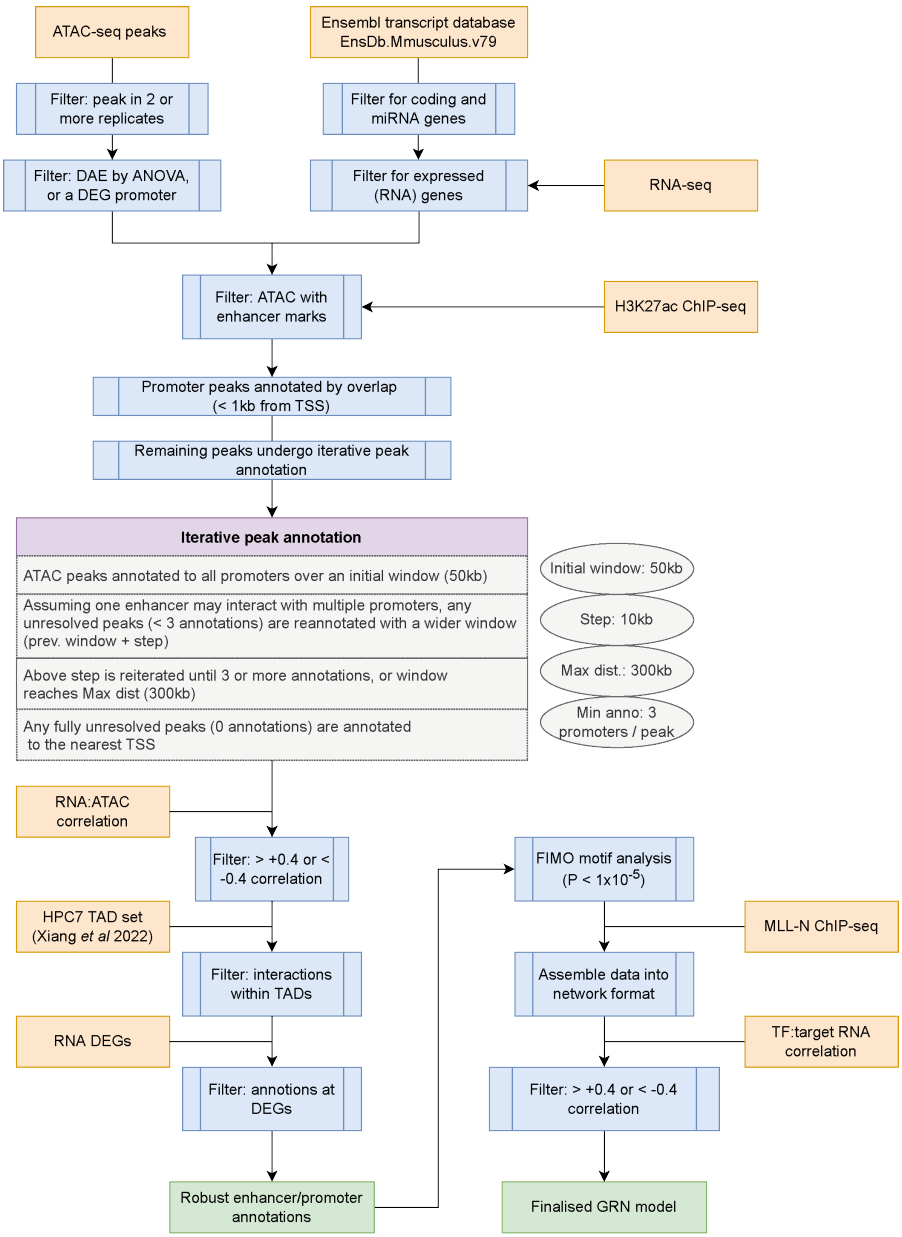


Figure A.10: Flow diagram explaining the processing of ATAC-seq, RNA-seq and ChIP-mentation data to generate the MLL-AF9 GMP GRN. Flow diagram schematic outlining the steps involved in the processing and annotating of ATAC-seq data, and the integration of gene expression and ChIP-mentation data to generate the MLL-AF9 GMP GRN. Input data coloured orange; data processing steps coloured blue; data outputs coloured green.

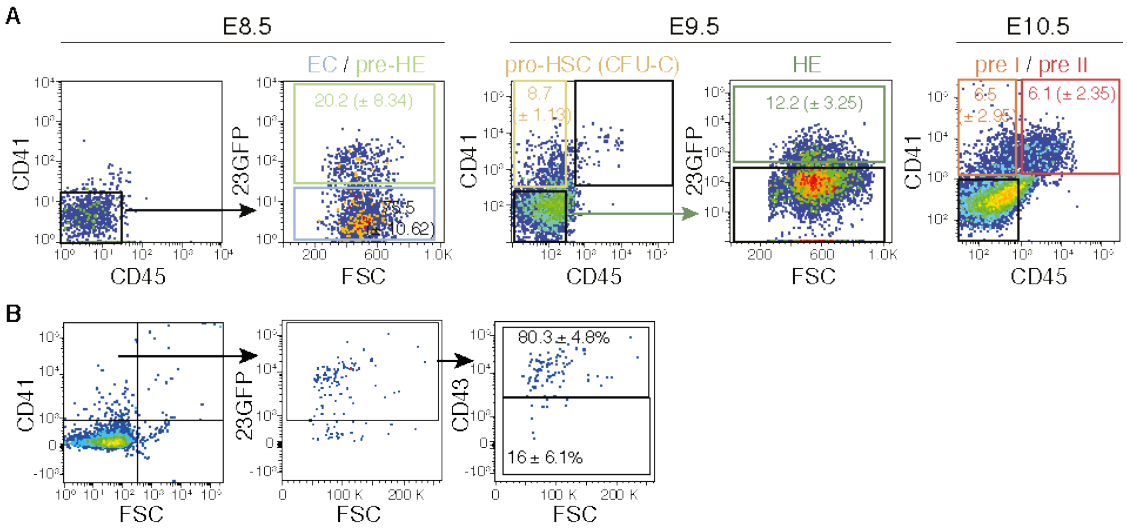


Figure A.11: Flow cytometry panels for EHT haematopoietic population isolation. (A) Representative flow cytometry panels for the sorting of EHT populations for RNA-seq and ATAC-seq analysis. Flow cytometry panels are gated on Ter119$^-$VE-Cad$^+$ live cells. **(B)** Representative flow cytometry panels analysing E9.5 PAS from 23GFP embryos, showing the percentage of CD43$^+$ cells in the Ter119$^-$VE-Cad$^+$CD41$^+$CD45$^-$ gate. *Flow cytometry analyses performed by Gemma Swiers and Emanuele Azzoni (de Bruijn lab).*

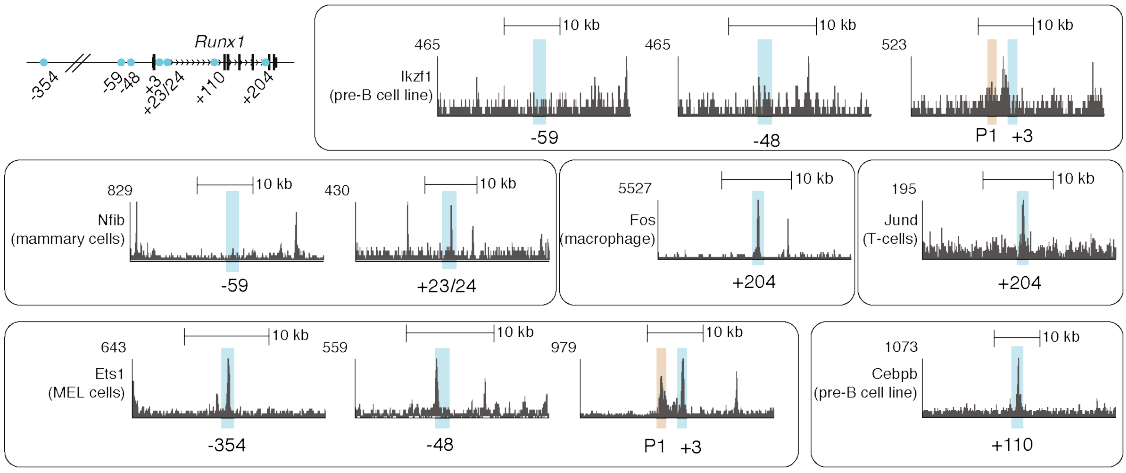


Figure A.12: Published ChIP-seq validation of TF motifs. ChIP-seq tracks over *Runx1* enhancers from reanalysed published ChIP-seq datasets. Schematic of *Runx1* locus and relevant enhancers shown. Tracks are scaled to maximum signal of the window.

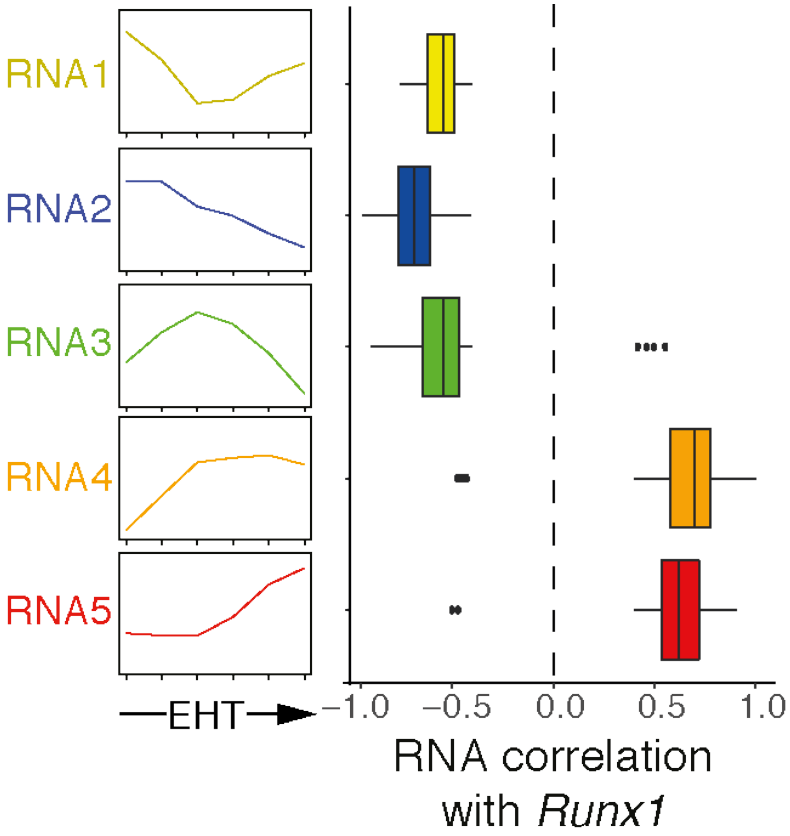


Figure A.13: *Runx1* correlation with predicted Runx1 targets, across EHT GRN modules. Boxplot summary of transcription correlation between *Runx1* expression and Runx1 target genes, stratified by RNA expression module (RNA1-5).

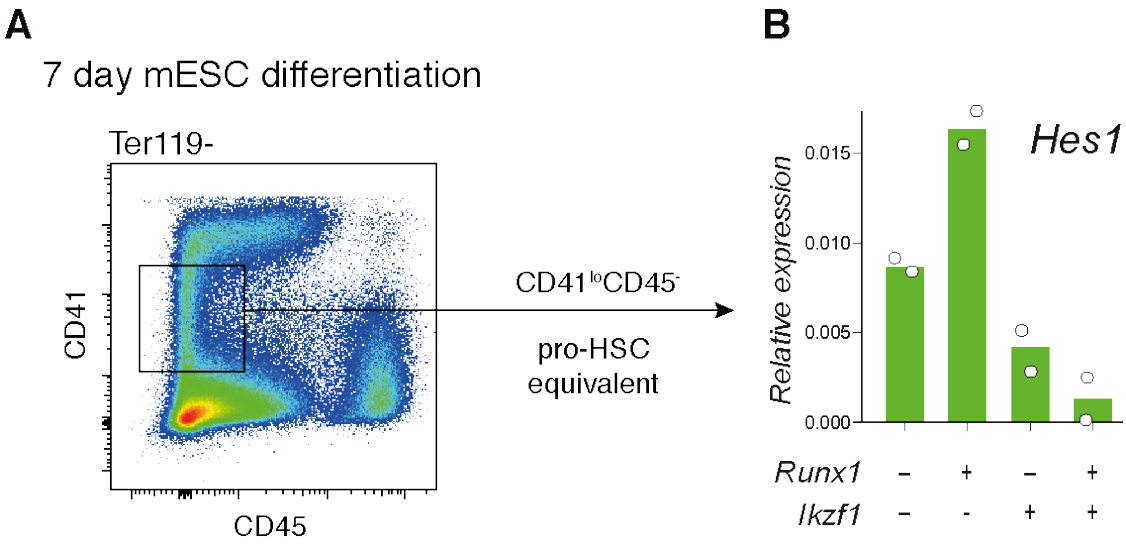


Figure A.14: Functional validation of the Runx1:Ikzf1:*Hes1* FFL. (A) Representative flow cytometry panel for analysis at day 7 of haematopoietic differentiation of mESC lines. Analyses were performed on Ter119^{-} live cells. **(B)** qRT-PCR assaying Hes1 expression at day 7 of mESC differentiation in Runx1-ERt2::Ikzf1$^{+/+}$ and Runx1-ERt2::Ikzf1$^{-/-}$ mESC lines, with or without +4-OHT. Expression normalized to Atp5a1 and Gapdh. N = 2 differentiations. Conditions are represented on the x axis as Runx1 and Ikzf1 status, where + indicates active, and - indicates inactive. Labels include Runx1- (-4-OHT), Runx1+ (+4-OHT), Ikzf1- (Runx1-ERt2::Ikzf1$^{-/-}$ cell line), and Ikzf1+ (Runx1-ERt2::Ikzf1$^{+/+}$ cell line). *Data in this figure was generated by Lucas Greder (de Bruijn lab)*

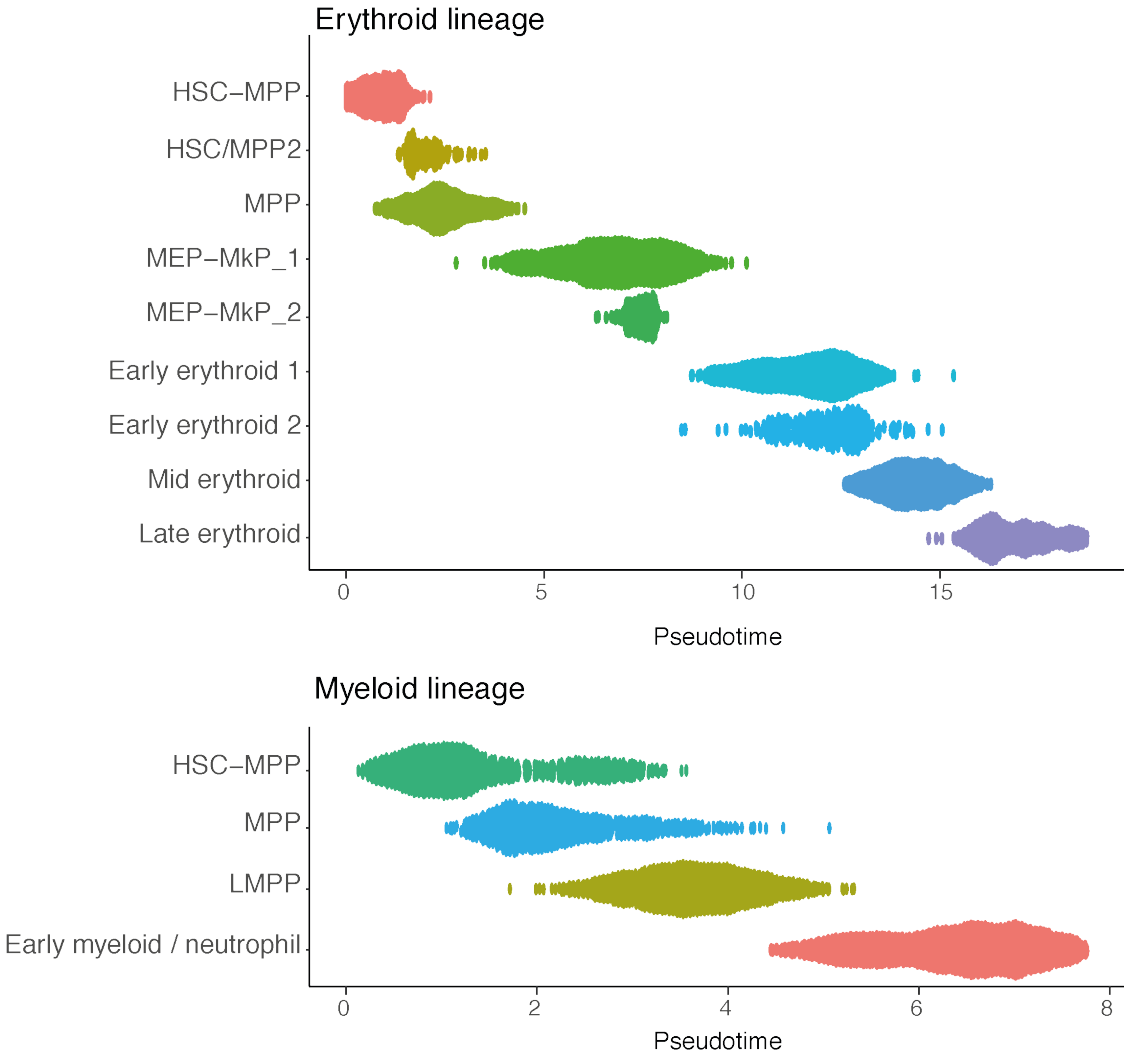


Figure A.15: Sc-multiome erythroid and myeloid (neutrophil) lineages. Erythropoiesis and Myelopoiesis assigned cells ordered by slingshot pseudotime, stratified by assigned cluster as in Figure 4.17.

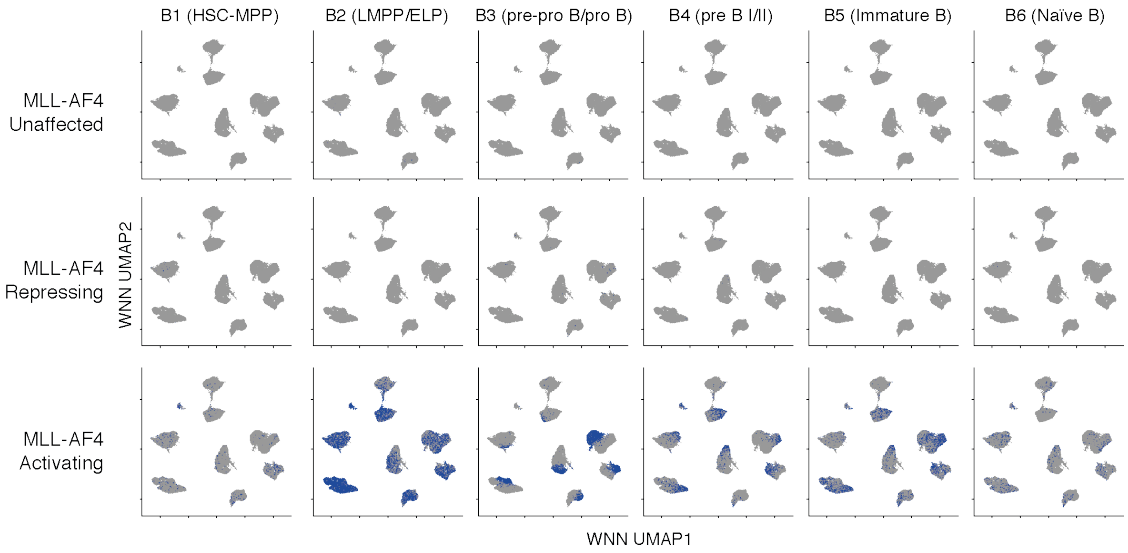


Figure A.16: MLL-AF4 GRN enrichment in normal FBM and FL. ALL blast UMAPs, annotated with cells passing AUCell thresholds marking enrichment for MLL-AF4 GRN regulated genes (MLL-AF4 activating or repressing) in each B lymphopoiesis signature as in Figure 4.18.

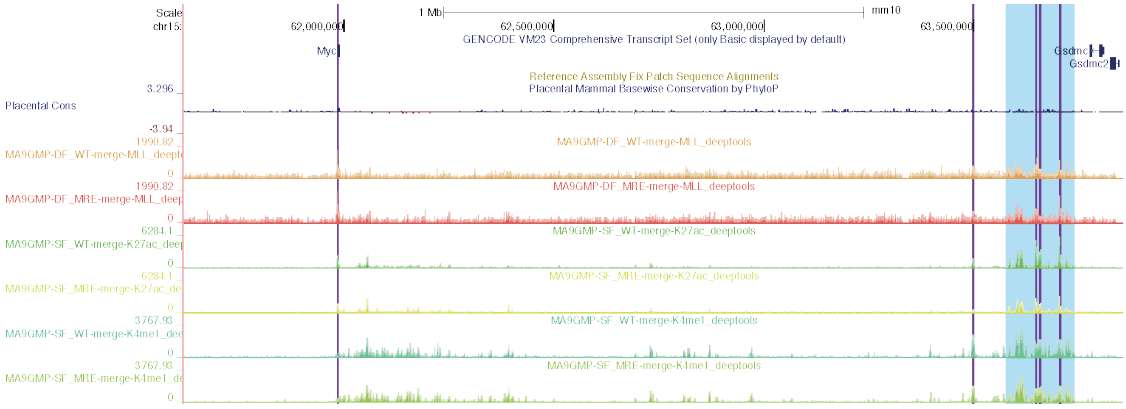


Figure A.17: *Myc* enhancer annotations in MLL-AF9 GMPs. UCSC tracks showing MLL-N, H3K27ac, and H3K4me1 ChIP-mentation tracks. Replicates were merged for visualisation. Blue highlight indicates 1.7 mb superenhancer, and purple highlights indicate enhancers elements annotated to *Myc*.

References

Agraz-Doblas, A., Bueno, C., Bashford-Rogers, R., Roy, A., Schneider, P., Bardini, M., Ballerini, P., Cazzaniga, G., Moreno, T., Revilla, C., Gut, M., Valsecchi, M. G., Roberts, I., Pieters, R., De Lorenzo, P., Varela, I., Menendez, P., & Stam, R. W. (2019). Unraveling the cellular origin and clinical prognostic markers of infant B-cell acute lymphoblastic leukemia using genome-wide analysis. *Haematologica*, *104*(6), 1176–1188. https://doi.org/10.3324/haematol.2018.206375

Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., van den Oord, J., Atak, Z. K., Wouters, J., & Aerts, S. (2017). SCENIC: Single-cell regulatory network inference and clustering. *Nature Methods*, *14*(11), 1083–1086. https://doi.org/10.1038/nmeth.4463

Allen, B. L., & Taatjes, D. J. (2015). The Mediator complex: A central integrator of transcription. *Nature Reviews Molecular Cell Biology*, *16*(3), 155–166. https://doi.org/10.1038/nrm3951

Alomairi, J., Molitor, A. M., Sadouni, N., Hussain, S., Torres, M., Saadi, W., Dao, L. T. M., Charbonnier, G., Santiago-Algarra, D., Andrau, J. C., Puthier, D., Sexton, T., & Spicuglia, S. (2020). Integration of high-throughput reporter assays identify a critical enhancer of the Ikzf1 gene. *PLoS ONE*, *15*(5), e0233191. https://doi.org/10.1371/journal.pone.0233191

Amemiya, H. M., Kundaje, A., & Boyle, A. P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific Reports*, *9*(1), 9354. https://doi.org/10.1038/s41598-019-45839-z

Andersson, A. K., Ma, J., Wang, J., Chen, X., Gedman, A. L., Dang, J., Nakitandwe, J., Holmfeldt, L., Parker, M., Easton, J., Huether, R., Kriwacki, R., Rusch, M., Wu, G., Li, Y., Mulder, H., Raimondi, S., Pounds, S., Kang, G., ... St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project. (2015). The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nature Genetics*, *47*(4), 330–337. https://doi.org/10.1038/ng.3230

Aratani, Y. (2018). Myeloperoxidase: Its role for host defense, inflammation, and neutrophil function. *Archives of Biochemistry and Biophysics*, *640*, 47–52. https://doi.org/10.1016/j.abb.2018.01.004

Assi, S. A., Imperato, M. R., Coleman, D. J. L., Pickin, A., Potluri, S., Ptasinska, A., Chin, P. S., Blair, H., Cauchy, P., James, S. R., Zacarias-Cabeza, J., Gilding, L. N.,

Beggs, A., Clokie, S., Loke, J. C., Jenkin, P., Uddin, A., Delwel, R., Richards, S. J., ... Bonifer, C. (2019). Subtype-specific regulatory network rewiring in acute myeloid leukemia. *Nature Genetics*, *51*(1), 151–162. https://doi.org/10.1038/s41588-018-0270-1

Baek, H., Kang, Y., & Roeder, R. (2006). Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. *The Journal of biological chemistry*, *281*(22). https://doi.org/10.1074/jbc.M601983200

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research*, *43*(W1), W39–W49. https://doi.org/10.1093/nar/gkv416

Ballabio, E., & Milne, T. A. (2012). Molecular and Epigenetic Mechanisms of MLL in Human Leukemogenesis. *Cancers*, *4*(3), 904–944. https://doi.org/10.3390/cancers4030904

Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, *27*(2 Pt 1), 299–308. https://doi.org/10.1016/0092-8674(81)90413-x

Baniahmad, A., Muller, M., Steiner, C., & Renkawitz, R. (1987). Activity of two different silencer elements of the chicken lysozyme gene can be compensated by enhancer elements. *The EMBO journal*, *6*(8), 2297–2303. https://doi.org/10.1002/j.1460-2075.1987.tb02504.x

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395. https://doi.org/10.1038/cr.2011.22

Bardini, M., Galbiati, M., Lettieri, A., Bungaro, S., Gorletta, T. A., Biondi, A., & Cazzaniga, G. (2011). Implementation of array based whole-genome high-resolution technologies confirms the absence of secondary copy-number alterations in MLL-AF4-positive infant ALL patients. *Leukemia*, *25*(1), 175–178. https://doi.org/10.1038/leu.2010.232

Bardini, M., Spinelli, R., Bungaro, S., Mangano, E., Corral, L., Cifola, I., Fazio, G., Giordan, M., Basso, G., De Rossi, G., Biondi, A., Battaglia, C., & Cazzaniga, G. (2010). DNA copy-number abnormalities do not occur in infant ALL with t(4;11)/MLL-AF4. *Leukemia*, *24*(1), 169–176. https://doi.org/10.1038/leu.2009.203

Baron, C. S., Kester, L., Klaus, A., Boisset, J.-C., Thambyrajah, R., Yvernogeau, L., Kouskoff, V., Lacaud, G., van Oudenaarden, A., & Robin, C. (2018). Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the

aorta. *Nature Communications*, *9*(1), 2517. https://doi.org/10.1038/s41467-018-04893-3

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., & Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, *129*(4), 823–837. https://doi.org/10.1016/j.cell.2007.05.009

Basu, S., Mehreja, R., Thiberge, S., Chen, M.-T., & Weiss, R. (2004). Spatiotemporal control of gene expression with pulse-generating networks. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(17), 6355–6360. https://doi.org/10.1073/pnas.0307571101

Batta, K., Florkowska, M., Kouskoff, V., & Lacaud, G. (2014). Direct reprogramming of murine fibroblasts to hematopoietic progenitor cells. *Cell Reports*, *9*(5), 1871–1884. https://doi.org/10.1016/j.celrep.2014.11.002

Bavelas, A. (1950). Communication Patterns in Task-Oriented Groups. *The Journal of the Acoustical Society of America*, *22*(6), 725–730. https://doi.org/10.1121/1.1906679

Baxter, E., Windloch, K., Gannon, F., & Lee, J. S. (2014). Epigenetic regulation in cancer progression. *Cell & Bioscience*, *4*, 45. https://doi.org/10.1186/2045-3701-4-45

Baxter, R. C. (2013). Insulin-like growth factor binding protein-3 (IGFBP-3): Novel ligands mediate unexpected functions. *Journal of Cell Communication and Signaling*, *7*(3), 179–189. https://doi.org/10.1007/s12079-013-0203-9

Beagan, J. A., & Phillips-Cremins, J. E. (2020). On the existence and functionality of topologically associating domains. *Nature Genetics*, *52*(1), 8–16. https://doi.org/10.1038/s41588-019-0561-1

Bee, T., Swiers, G., Muroi, S., Pozner, A., Nottingham, W., Santos, A. C., Li, P.-S., Taniuchi, I., & de Bruijn, M. F. T. R. (2010). Nonredundant roles for Runx1 alternative promoters reflect their activity at discrete stages of developmental hematopoiesis. *Blood*, *115*(15), 3042–3050. https://doi.org/10.1182/blood-2009-08-238626

Behan, F. M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C. M., Migliardi, G., Santos, R., Rao, Y., Sassi, F., Pinnelli, M., Ansari, R., Harper, S., Jackson, D. A., McRae, R., Pooley, R., Wilkinson, P., Meer, D. v. d., Dow, D., Buser-Doepner, C., … Garnett, M. J. (2019). Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature*, *568*(7753), 511–516. https://doi.org/10.1038/s41586-019-1103-9

Benito, J. M., Godfrey, L., Kojima, K., Hogdal, L., Wunderlich, M., Geng, H., Marzo, I., Harutyunyan, K. G., Golfman, L., North, P., Kerry, J., Ballabio, E., Chonghaile, T. N., Gonzalo, O., Qiu, Y., Jeremias, I., Debose, L., O'Brien, E., Ma, H., ... Konopleva, M. (2015). MLL-Rearranged Acute Lymphoblastic Leukemias Activate BCL-2 through H3K79 Methylation and Are Sensitive to the BCL-2-Specific Antagonist ABT-199. *Cell Reports*, *13*(12), 2715–2727. https://doi.org/10.1016/j.celrep.2015.12.003

Bergiers, I., Andrews, T., Vargel Bölükbaşı, Ö., Buness, A., Janosz, E., Lopez-Anguita, N., Ganter, K., Kosim, K., Celen, C., Itır Perçin, G., Collier, P., Baying, B., Benes, V., Hemberg, M., & Lancrin, C. (2018). Single-cell transcriptomics reveals a new dynamical function of transcription factors during embryonic hematopoiesis (A. J. Wagers, Ed.). *eLife*, *7*, e29312. https://doi.org/10.7554/eLife.29312

Bernt, K. M., Zhu, N., Sinha, A. U., Vempati, S., Faber, J., Krivtsov, A. V., Feng, Z., Punt, N., Daigle, A., Bullinger, L., Pollock, R. M., Richon, V. M., Kung, A. L., & Armstrong, S. A. (2011). MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell*, *20*(1), 66–78. https://doi.org/10.1016/j.ccr.2011.06.010

Bertrand, J. Y., Chi, N. C., Santoso, B., Teng, S., Stainier, D. Y. R., & Traver, D. (2010). Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature*, *464*(7285), 108–111. https://doi.org/10.1038/nature08738

Bhattacharjee, S., Renganaath, K., Mehrotra, R., & Mehrotra, S. (2013). Combinatorial Control of Gene Expression. *BioMed Research International*, *2013*, 407263. https://doi.org/10.1155/2013/407263

Birke, M., Schreiner, S., García-Cuéllar, M.-P., Mahr, K., Titgemeyer, F., & Slany, R. K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Research*, *30*(4), 958–965. https://doi.org/10.1093/nar/30.4.958

Bissonnette, R. P., Echeverri, F., Mahboubi, A., & Green, D. R. (1992). Apoptotic cell death induced by c-myc is inhibited by bcl-2. *Nature*, *359*(6395), 552–554. https://doi.org/10.1038/359552a0

Biswas, D., Milne, T. A., Basrur, V., Kim, J., Elenitoba-Johnson, K. S. J., Allis, C. D., & Roeder, R. G. (2011). Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(38), 15751–15756. https://doi.org/10.1073/pnas.1111498108

Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L. P., Ito, S., Cooper, S., Kondo, K., Koseki, Y., Ishikura, T., Long, H. K., Sheahan, T. W., Brockdorff, N., Kessler, B. M., Koseki, H., & Klose, R. J. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell*, *157*(6), 1445–1459. https://doi.org/10.1016/j.cell.2014.05.004

Boeva, V. (2016). Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells. *Frontiers in Genetics*, *7*.

Bohne, A., Schlee, C., Mossner, M., Thibaut, J., Heesch, S., Thiel, E., Hofmann, W.-K., & Baldus, C. D. (2009). Epigenetic control of differential expression of specific ERG isoforms in acute T-lymphoblastic leukemia. *Leukemia Research*, *33*(6), 817–822. https://doi.org/10.1016/j.leukres.2008.11.012

Boisset, J.-C., van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E., & Robin, C. (2010). In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature*, *464*(7285), 116–120. https://doi.org/10.1038/nature08764

Boller, S., & Grosschedl, R. (2014). The regulatory network of B-cell differentiation: A focused view of early B-cell factor 1 function [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/imr.12206]. *Immunological Reviews*, *261*(1), 102–115. https://doi.org/10.1111/imr.12206

Braccioli, L., & de Wit, E. (2019). CTCF: A Swiss-army knife for genome organization and transcription regulation. *Essays in Biochemistry*, *63*(1), 157–165. https://doi.org/10.1042/EBC20180069

Brand, A. H., Breeden, L., Abraham, J., Sternglanz, R., & Nasmyth, K. (1985). Characterization of a "silencer" in yeast: A DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell*, *41*(1), 41–48. https://doi.org/10.1016/0092-8674(85)90059-5

Brand, A. H., Micklem, G., & Nasmyth, K. (1987). A yeast silencer contains sequences that can promote autonomous plasmid replication and transcriptional activation. *Cell*, *51*(5), 709–719. https://doi.org/10.1016/0092-8674(87)90094-8

Brent, R., & Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell*, *43*(3 Pt 2), 729–736. https://doi.org/10.1016/0092-8674(85)90246-6

Britten, R. J., & Davidson, E. H. (1969). Gene regulation for higher cells: A theory. *Science (New York, N.Y.)*, *165*(3891), 349–357. https://doi.org/10.1126/science.165.3891.349

Buenrostro, J., Wu, B., Chang, H., & Greenleaf, W. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, *109*, 21.29.1–21.29.9. https://doi.org/10.1002/0471142727.mb2129s109

Burns, C. E., Traver, D., Mayhall, E., Shepard, J. L., & Zon, L. I. (2005). Hematopoietic stem cell fate is established by the Notch–Runx pathway. *Genes & Development*, *19*(19), 2331–2342. https://doi.org/10.1101/gad.1337005

Byrne, A. B., Weirauch, M. T., Wong, V., Koeva, M., Dixon, S. J., Stuart, J. M., & Roy, P. J. (2007). A global analysis of genetic interactions in Caenorhabditis elegans. *Journal of Biology*, *6*(3), 8. https://doi.org/10.1186/jbiol58

Cai, Z., de Bruijn, M., Ma, X., Dortland, B., Luteijn, T., Downing, R. J., & Dzierzak, E. (2000). Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. *Immunity*, *13*(4), 423–431. https://doi.org/10.1016/s1074-7613(00)00042-x

Canon, J., & Banerjee, U. (2003). In vivo analysis of a developmental circuit for direct transcriptional activation and repression in the same cell by a Runx protein. *Genes & Development*, *17*(7), 838–843. https://doi.org/10.1101/gad.1064803

Carpenter, L., Malladi, R., Yang, C.-T., French, A., Pilkington, K. J., Forsey, R. W., Sloane-Stanley, J., Silk, K. M., Davies, T. J., Fairchild, P. J., Enver, T., & Watt, S. M. (2011). Human induced pluripotent stem cells are capable of B-cell lymphopoiesis. *Blood*, *117*(15), 4008–4011. https://doi.org/10.1182/blood-2010-08-299941

Castro, P. G., Bomken, S., Seslija, L., Stam, R., Latif, E. S., Batey, M. A., McNeill, H., Pieters, R., Vormoor, J., Greil, J., & Heidenreich, O. (2010). ANGIOPOIETIN1 - a Novel Factor Implicated In MLL-Rearranged Acute Lymphoblastic Leukemia and Regulated In a Fusion Gene-Dependent Manner. *Blood*, *116*(21), 66. https://doi.org/10.1182/blood.V116.21.66.66

Cauchy, P., James, S. R., Zacarias-Cabeza, J., Ptasinska, A., Imperato, M. R., Assi, S. A., Piper, J., Canestraro, M., Hoogenkamp, M., Raghavan, M., Loke, J., Akiki, S., Clokie, S. J., Richards, S. J., Westhead, D. R., Griffiths, M. J., Ott, S., Bonifer, C., & Cockerill, P. N. (2015). Chronic FLT3-ITD Signaling in Acute Myeloid Leukemia Is Connected to a Specific Chromatin Signature. *Cell Reports*, *12*(5), 821–836. https://doi.org/10.1016/j.celrep.2015.06.069

Chaudhry, P., Singh, M., Triche, T. J., Guzman, M., & Merchant, A. A. (2017). GLI3 repressor determines Hedgehog pathway activation and is required for response to SMO antagonist glasdegib in AML. *Blood*, *129*(26), 3465–3475. https://doi.org/10.1182/blood-2016-05-718585

Chaudhuri, A. A., So, A. Y.-L., Mehta, A., Minisandram, A., Sinha, N., Jonsson, V. D., Rao, D. S., O'Connell, R. M., & Baltimore, D. (2012). Oncomir miR-125b regulates hematopoiesis by targeting the gene Lin28A. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(11), 4233–4238. https://doi.org/10.1073/pnas.1200677109

Chen, B., Brinkmann, K., Chen, Z., Pak, C. W., Liao, Y., Shi, S., Henry, L., Grishin, N. V., Bogdan, S., & Rosen, M. K. (2014). The WAVE Regulatory Complex Links Diverse Receptors to the Actin Cytoskeleton. *Cell*, *156*(1), 195–207. https://doi.org/10.1016/j.cell.2013.11.048

Chen, M. J., Yokomizo, T., Zeigler, B., Dzierzak, E., & Speck, N. A. (2009). Runx1 is required for the endothelial to hematopoietic cell transition but not thereafter. *Nature*, *457*(7231), 887–891. https://doi.org/10.1038/nature07619

Chen, S., & Mar, J. C. (2018). Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics*, *19*(1), 232. https://doi.org/10.1186/s12859-018-2217-z

Chen, Z., Borek, D., Padrick, S. B., Gomez, T. S., Metlagel, Z., Ismail, A., Umetani, J., Billadeau, D. D., Otwinowski, Z., & Rosen, M. K. (2010). Structure and Control of the Actin Regulatory WAVE Complex. *Nature*, *468*(7323), 533–538. https://doi.org/10.1038/nature09623

Cheng, C.-K., Wong, T. H. Y., Wan, T. S. K., Wang, A. Z., Chan, N. P. H., Chan, N. C. N., Li, C.-K., & Ng, M. H. L. (2018). RUNX1 upregulation via disruption of long-range transcriptional control by a novel t(5;21)(q13;q22) translocation in acute myeloid leukemia. *Molecular Cancer*, *17*(1), 133. https://doi.org/10.1186/s12943-018-0881-2

Chepelev, I., Wei, G., Wangsa, D., Tang, Q., & Zhao, K. (2012). Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Research*, *22*(3), 490–503. https://doi.org/10.1038/cr.2012.15

Cherry, E. M., Abbott, D., Amaya, M., McMahon, C., Schwartz, M., Rosser, J., Sato, A., Schowinsky, J., Inguva, A., Minhajuddin, M., Pei, S., Stevens, B., Winters, A., Jordan, C. T., Smith, C., Gutman, J. A., & Pollyea, D. A. (2021). Venetoclax and azacitidine compared with induction chemotherapy for newly

diagnosed patients with acute myeloid leukemia. *Blood Advances*, *5*(24), 5565–5573. https://doi.org/10.1182/bloodadvances.2021005538

Choi, A., Illendula, A., Pulikkan, J. A., Roderick, J. E., Tesell, J., Yu, J., Hermance, N., Zhu, L. J., Castilla, L. H., Bushweller, J. H., & Kelliher, M. A. (2017). RUNX1 is required for oncogenic Myb and Myc enhancer activity in T-cell acute lymphoblastic leukemia. *Blood*, *130*(15), 1722–1733. https://doi.org/10.1182/blood-2017-03-775536

Chow, S., Tang, K., Al-Abri, M., Hall, V., Tremblay-Lemay, R., Rashedi, I., Tsui, H., & Chan, S. M. (2021). RUNX1 mutations correlate with response to venetoclax combination therapies in relapsed/refractory acute myeloid leukemia. *Leukemia Research*, *111*, 106735. https://doi.org/10.1016/j.leukres.2021.106735

Ciofani, M., Madar, A., Galan, C., Sellars, M., Mace, K., Pauli, F., Agarwal, A., Huang, W., Parkhurst, C. N., Muratet, M., Newberry, K. M., Meadows, S., Greenfield, A., Yang, Y., Jain, P., Kirigin, F. F., Birchmeier, C., Wagner, E. F., Murphy, K. M., . . . Littman, D. R. (2012). A validated regulatory network for Th17 cell specification. *Cell*, *151*(2), 289–303. https://doi.org/10.1016/j.cell.2012.09.016

Clarke, R. L., Yzaguirre, A. D., Yashiro-Ohtani, Y., Bondue, A., Blanpain, C., Pear, W. S., Speck, N. A., & Keller, G. (2013). The expression of Sox17 identifies and regulates haemogenic endothelium. *Nature Cell Biology*, *15*(5), 502–510. https://doi.org/10.1038/ncb2724

Cornejo-Páramo, P., Roper, K., Degnan, S. M., Degnan, B. M., & Wong, E. S. (2022). Distal regulation, silencers, and a shared combinatorial syntax are hallmarks of animal embryogenesis. *Genome Research*, *32*(3), 474–487. https://doi.org/10.1101/gr.275864.121

Cramer, P. (2019). Organization and regulation of gene transcription. *Nature*, *573*(7772), 45–54. https://doi.org/10.1038/s41586-019-1517-4

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21931–21936. https://doi.org/10.1073/pnas.1016071107

Crosse, E. I., Gordon-Keylock, S., Rybtsov, S., Binagui-Casas, A., Felchle, H., Nnadi, N. C., Kirschner, K., Chandra, T., Tamagno, S., Webb, D. J., Rossi, F., Anderson, R. A., & Medvinsky, A. (2020). Multi-layered Spatial Transcriptomics Identify Secretory Factors Promoting Human Hematopoietic Stem Cell Development. *Cell Stem Cell*, *27*(5), 822–839.e8. https://doi.org/10.1016/j.stem.2020.08.004

Crump, N. T., Ballabio, E., Godfrey, L., Thorne, R., Repapi, E., Kerry, J., Tapia, M., Hua, P., Lagerholm, C., Filippakopoulos, P., Davies, J. O. J., & Milne, T. A. (2021). BET inhibition disrupts transcription but retains enhancer-promoter contact. *Nature Communications*, *12*(1), 223. https://doi.org/10.1038/s41467-020-20400-z

Crump, N. T., & Milne, T. A. (2019). Why are so many MLL lysine methyltransferases required for normal mammalian development? *Cellular and Molecular Life Sciences*, *76*(15), 2885–2898. https://doi.org/10.1007/s00018-019-03143-z

Crump, N. T., Smith, A., Godfrey, L., Jackson, N., Rice, S., Kim, J., Basrur, V., Fermin, D., Elenitoba-Johnson, K., Roeder, R. G., Allis, C. D., Roberts, I., Roy, A., & Milne, T. A. (2022). PAF1 and FACT cooperate with MLL-AF4 to drive enhancer activity in leukemia [Pages: 2022.05.27.493676 Section: New Results]. https://doi.org/10.1101/2022.05.27.493676

Cuvertino, S., Filiciotto, G., Masurekar, A., Saha, V., Lacaud, G., & Kouskoff, V. (2016). SOX7 promotes the maintenance and proliferation of B cell precursor acute lymphoblastic cells. *Oncotarget*, *8*(39), 64974–64983. https://doi.org/10.18632/oncotarget.10472

Czabotar, P. E., Lessene, G., Strasser, A., & Adams, J. M. (2014). Control of apoptosis by the BCL-2 protein family: Implications for physiology and therapy. *Nature Reviews Molecular Cell Biology*, *15*(1), 49–63. https://doi.org/10.1038/nrm3722

Dai, P., Akimaru, H., Tanaka, Y., Hou, D. X., Yasukawa, T., Kanei-Ishii, C., Takahashi, T., & Ishii, S. (1996). CBP as a transcriptional coactivator of c-Myb. *Genes & Development*, *10*(5), 528–540. https://doi.org/10.1101/gad.10.5.528

Dang, C. V. (2012). MYC on the Path to Cancer. *Cell*, *149*(1), 22–35. https://doi.org/10.1016/j.cell.2012.03.003

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust, A. G., Pan, Z. j., Schilstra, M. J., Clarke, P. J. C., … Bolouri, H. (2002). A genomic regulatory network for development. *Science (New York, N.Y.)*, *295*(5560), 1669–1678. https://doi.org/10.1126/science.1069883

Dawson, M. A., Prinjha, R. K., Dittmann, A., Giotopoulos, G., Bantscheff, M., Chan, W.-I., Robson, S. C., Chung, C.-w., Hopf, C., Savitski, M. M., Huthmacher, C., Gudgin, E., Lugo, D., Beinke, S., Chapman, T. D., Roberts, E. J., Soden, P. E., Auger, K. R., Mirguet, O., … Kouzarides, T. (2011). Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature*, *478*(7370), 529–533. https://doi.org/10.1038/nature10509

de Bruijn, M. F., Speck, N. A., Peeters, M. C., & Dzierzak, E. (2000). Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *The EMBO journal*, *19*(11), 2465–2474. https://doi.org/10.1093/emboj/19.11.2465

de Matos Simoes, R., Dehmer, M., & Emmert-Streib, F. (2013). B-cell lymphoma gene regulatory networks: Biological consistency among inference methods. *Frontiers in Genetics*, *4*, 281. https://doi.org/10.3389/fgene.2013.00281

Dey, A., Chitsaz, F., Abbasi, A., Misteli, T., & Ozato, K. (2003). The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(15), 8758–8763. https://doi.org/10.1073/pnas.1433065100

Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K., & Zhou, M. M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature*, *399*(6735), 491–496. https://doi.org/10.1038/20974

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R., Adamson, B., Norman, T. M., Lander, E. S., Weissman, J. S., Friedman, N., & Regev, A. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*, *167*(7), 1853–1866.e17. https://doi.org/10.1016/j.cell.2016.11.038

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–380. https://doi.org/10.1038/nature11082

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, *34*(2), 184–191. https://doi.org/10.1038/nbt.3437

Dorantes-Acosta, E., & Pelayo, R. (2012). Lineage Switching in Acute Leukemias: A Consequence of Stem Cell Plasticity? *Bone Marrow Research*, *2012*, 406796. https://doi.org/10.1155/2012/406796

Dunn, R., Dudbridge, F., & Sanderson, C. M. (2005). The Use of Edge-Betweenness Clustering to Investigate Biological Function in Protein Interaction Networks. *BMC Bioinformatics*, *6*(1), 39. https://doi.org/10.1186/1471-2105-6-39

Eguchi-Ishimae, M., Eguchi, M., Wu, Z., Ming, W., Iwabuki, H., Inukai, T., Sugita, K., & Ishii, E. (2014). HMGA2 As a Potential Molecular Target in MLL-AF4 Positive Infant Acute Lymphoblastic Leukemia. *Blood*, *124*(21), 2244. https://doi.org/10.1182/blood.V124.21.2244.2244

Eichenfield, D. Z., Troutman, T. D., Link, V. M., Lam, M. T., Cho, H., Gosselin, D., Spann, N. J., Lesch, H. P., Tao, J., Muto, J., Gallo, R. L., Evans, R. M., & Glass, C. K. (2016). Tissue damage drives co-localization of NF-ϰB, Smad3, and Nrf2 to direct Rev-erb sensitive wound repair in mouse macrophages. *eLife*, *5*, e13024. https://doi.org/10.7554/eLife.13024

El Amri, M., Fitzgerald, U., & Schlosser, G. (2018). MARCKS and MARCKS-like proteins in development and regeneration. *Journal of Biomedical Science*, *25*, 43. https://doi.org/10.1186/s12929-018-0445-1

Eliades, A., Wareing, S., Marinopoulou, E., Fadlullah, M. Z., Patel, R., Grabarek, J. B., Plusa, B., Lacaud, G., & Kouskoff, V. (2016). The Hemogenic Competence of Endothelial Progenitors Is Restricted by Runx1 Silencing during Embryonic Development. *Cell Reports*, *15*(10), 2185–2199. https://doi.org/10.1016/j.celrep.2016.05.001

Emmert-Streib, F., De Matos Simoes, R., Mullan, P., Haibe-Kains, B., & Dehmer, M. (2014). The gene regulatory network for breast cancer: Integrated regulatory landscape of cancer hallmarks. *Frontiers in Genetics*, *5*, 15. https://doi.org/10.3389/fgene.2014.00015

Erman, B., Cortes, M., Nikolajczyk, B. S., Speck, N. A., & Sen, R. (1998). ETS-core binding factor: A common composite motif in antigen receptor gene enhancers. *Molecular and Cellular Biology*, *18*(3), 1322–1330. https://doi.org/10.1128/MCB.18.3.1322

Fairlie, W. D., & Lee, E. F. (2021). Co-Operativity between MYC and BCL-2 Pro-Survival Proteins in Cancer. *International Journal of Molecular Sciences*, *22*(6), 2841. https://doi.org/10.3390/ijms22062841

Fanidi, A., Harrington, E. A., & Evan, G. I. (1992). Cooperative interaction between c-myc and bcl-2 proto-oncogenes. *Nature*, *359*(6395), 554–556. https://doi.org/10.1038/359554a0

Farrugia, A. J., & Calvo, F. (2016). The Borg family of Cdc42 effector proteins Cdc42EP1–5. *Biochemical Society Transactions*, *44*(6), 1709–1716. https://doi.org/10.1042/BST20160219

Ficara, F., Murphy, M. J., Lin, M., & Cleary, M. L. (2008). Pbx1 regulates self-renewal of long-term hematopoietic stem cells by maintaining their quiescence. *Cell Stem Cell*, *2*(5), 484–496. https://doi.org/10.1016/j.stem.2008.03.004

Fitch, S. R., Kapeni, C., Tsitsopoulou, A., Wilson, N. K., Göttgens, B., de Bruijn, M. F., & Ottersbach, K. (2020). Gata3 targets Runx1 in the embryonic haematopoietic stem cell niche. *IUBMB life*, *72*(1), 45–52. https://doi.org/10.1002/iub.2184

Ford, A. M., Ridge, S. A., Cabrera, M. E., Mahmoud, H., Steel, C. M., Chan, L. C., & Greaves, M. (1993). In utero rearrangements in the trithorax-related oncogene in infant leukaemias. *Nature*, *363*(6427), 358–360. https://doi.org/10.1038/363358a0

Frame, J. M., McGrath, K. E., & Palis, J. (2013). Erythro-myeloid progenitors: "Definitive" hematopoiesis in the conceptus prior to the emergence of hematopoietic stem cells. *Blood Cells, Molecules, and Diseases*, *51*(4), 220–225. https://doi.org/10.1016/j.bcmd.2013.09.006

Freedman, L. P., Luisi, B. F., Korszun, Z. R., Basavappa, R., Sigler, P. B., & Yamamoto, K. R. (1988). The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. *Nature*, *334*(6182), 543–546. https://doi.org/10.1038/334543a0

Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, *40*(1), 35–41. https://doi.org/10.2307/3033543

Galic, Z., Kitchen, S. G., Kacena, A., Subramanian, A., Burke, B., Cortado, R., & Zack, J. A. (2006). T lineage differentiation from human embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(31), 11742–11747. https://doi.org/10.1073/pnas.0604244103

Gao, P., Chen, C., Howell, E. D., Li, Y., Tober, J., Uzun, Y., He, B., Gao, L., Zhu, Q., Siekmann, A. F., Speck, N. A., & Tan, K. (2020). Transcriptional regulatory network controlling the ontogeny of hematopoietic stem cells. *Genes & Development*, doi: 10.1101/gad.338202.120. https://doi.org/10.1101/gad.338202.120

Gardner, R., Wu, D., Cherian, S., Fang, M., Hanafi, L.-A., Finney, O., Smithers, H., Jensen, M. C., Riddell, S. R., Maloney, D. G., & Turtle, C. J. (2016). Acquisition of a CD19-negative myeloid phenotype allows immune escape of MLL-rearranged B-ALL from CD19 CAR-T-cell therapy. *Blood*, *127*(20), 2406–2410. https://doi.org/10.1182/blood-2015-08-665547

Gekas, C., Rhodes, K. E., Gereige, L. M., Helgadottir, H., Ferrari, R., Kurdistani, S. K., Montecino-Rodriguez, E., Bassel-Duby, R., Olson, E., Krivtsov, A. V., Armstrong, S., Orkin, S. H., Pellegrini, M., & Mikkola, H. K. A. (2009). Mef2C is a lineage-restricted target of Scl/Tal1 and regulates megakaryopoiesis and B-cell homeostasis. *Blood*, *113*(15), 3461–3471. https://doi.org/10.1182/blood-2008-07-167577

Geng, H., Brennan, S., Milne, T. A., Chen, W.-Y., Li, Y., Hurtz, C., Kweon, S.-M., Zickl, L., Shojaee, S., Neuberg, D., Huang, C., Biswas, D., Xin, Y., Racevskis, J., Ketterling, R. P., Luger, S. M., Lazarus, H., Tallman, M. S., Rowe, J. M., … Melnick, A. M. (2012). Integrative Epigenomic Analysis Identifies Biomarkers and Therapeutic Targets in Adult B-Acute Lymphoblastic Leukemia. *Cancer Discovery*, *2*(11), 1004–1023. https://doi.org/10.1158/2159-8290.CD-12-0208

Gentek, R., Ghigo, C., Hoeffel, G., Bulle, M. J., Msallam, R., Gautier, G., Launay, P., Chen, J., Ginhoux, F., & Bajénoff, M. (2018). Hemogenic Endothelial Fate Mapping Reveals Dual Developmental Origin of Mast Cells. *Immunity*, *48*(6), 1160–1171.e5. https://doi.org/10.1016/j.immuni.2018.04.025

Georgescu, C., Longabaugh, W. J. R., Scripture-Adams, D. D., David-Fung, E.-S., Yui, M. A., Zarnegar, M. A., Bolouri, H., & Rothenberg, E. V. (2008). A gene regulatory network armature for T lymphocyte specification. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(51), 20100–20105. https://doi.org/10.1073/pnas.0806501105

Georgopoulos, K. (2017). The making of a lymphocyte: The choice among disparate cell fates and the IKAROS enigma. *Genes & Development*, *31*(5), 439–450. https://doi.org/10.1101/gad.297002.117

Giese, K., Kingsley, C., Kirshner, J. R., & Grosschedl, R. (1995). Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes & Development*, *9*(8), 995–1008. https://doi.org/10.1101/gad.9.8.995

Gilan, O., Lam, E. Y. N., Becher, I., Lugo, D., Cannizzaro, E., Joberty, G., Ward, A., Wiese, M., Fong, C. Y., Ftouni, S., Tyler, D., Stanley, K., MacPherson, L., Weng, C.-F., Chan, Y.-C., Ghisi, M., Smil, D., Carpenter, C., Brown, P., … Dawson, M. A. (2016). Functional interdependence of BRD4 and DOT1L in MLL leukemia. *Nature Structural & Molecular Biology*, *23*(7), 673–681. https://doi.org/10.1038/nsmb.3249

Gilmour, J., Assi, S. A., Noailles, L., Lichtinger, M., Obier, N., & Bonifer, C. (2018). The Co-operation of RUNX1 with LDB1, CDK9 and BRD4 Drives Transcription Factor Complex Relocation During Haematopoietic Specification. *Scientific Reports*, *8*, 10410. https://doi.org/10.1038/s41598-018-28506-7

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 7821–7826. https://doi.org/10.1073/pnas.122653799

Godfrey, L., Crump, N. T., O'Byrne, S., Lau, I.-J., Rice, S., Harman, J. R., Jackson, T., Elliott, N., Buck, G., Connor, C., Thorne, R., Knapp, D. J. H. F., Heidenreich, O., Vyas, P., Menendez, P., Inglott, S., Ancliff, P., Geng, H., Roberts, I., … Milne, T. A. (2021). H3K79me2/3 controls enhancer–promoter interactions and activation of the pan-cancer stem cell marker PROM1 /CD133 in MLL-AF4 leukemia cells. *Leukemia*, *35*(1), 90–106. https://doi.org/10.1038/s41375-020-0808-y

Godfrey, L., Crump, N. T., Thorne, R., Lau, I.-J., Repapi, E., Dimou, D., Smith, A. L., Harman, J. R., Telenius, J. M., Oudelaar, A. M., Downes, D. J., Vyas, P., Hughes, J. R., & Milne, T. A. (2019). DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. *Nature Communications*, *10*(1), 2803. https://doi.org/10.1038/s41467-019-10844-3

Godfrey, L., Kerry, J., Thorne, R., Repapi, E., Davies, J. O., Tapia, M., Ballabio, E., Hughes, J. R., Geng, H., Konopleva, M., & Milne, T. A. (2017). MLL-AF4 binds directly to a BCL-2 specific enhancer and modulates H3K27 acetylation. *Experimental Hematology*, *47*, 64–75. https://doi.org/10.1016/j.exphem.2016.11.003

Goentoro, L., Shoval, O., Kirschner, M., & Alon, U. (2009). The incoherent feedforward loop can provide fold-change detection in gene regulation. *Molecular cell*, *36*(5), 894–899. https://doi.org/10.1016/j.molcel.2009.11.018

Goetz, T. L., Gu, T. L., Speck, N. A., & Graves, B. J. (2000). Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. *Molecular and Cellular Biology*, *20*(1), 81–90. https://doi.org/10.1128/MCB.20.1.81-90.2000

Gomez Perdiguero, E., Klapproth, K., Schulz, C., Busch, K., Azzoni, E., Crozet, L., Garner, H., Trouillet, C., de Bruijn, M. F., Geissmann, F., & Rodewald, H.-R. (2015). Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature*, *518*(7540), 547–551. https://doi.org/10.1038/nature13989

González-Blas, C. B., Winter, S. D., Hulselmans, G., Hecker, N., Matetovici, I., Christiaens, V., Poovathingal, S., Wouters, J., Aibar, S., & Aerts, S. (2022). SCENIC+: Single-cell multiomic inference of enhancers and gene regulatory networks [Pages: 2022.08.19.504505 Section: New Results]. https://doi.org/10.1101/2022.08.19.504505

Goode, D. K., Obier, N., Vijayabaskar, M. S., Lie-A-Ling, M., Lilly, A. J., Hannah, R., Lichtinger, M., Batta, K., Florkowska, M., Patel, R., Challinor, M., Wallace, K., Gilmour, J., Assi, S. A., Cauchy, P., Hoogenkamp, M., Westhead, D. R., Lacaud, G., Kouskoff, V., ... Bonifer, C. (2016). Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. *Developmental Cell*, *36*(5), 572–587. https://doi.org/10.1016/j.devcel.2016.01.024

Goumans, M.-J., Valdimarsdottir, G., Itoh, S., Rosendahl, A., Sideras, P., & ten Dijke, P. (2002). Balancing the activation state of the endothelium via two distinct TGF-beta type I receptors. *The EMBO journal*, *21*(7), 1743–1753. https://doi.org/10.1093/emboj/21.7.1743

Granja, J. M., Corces, M. R., Pierce, S. E., Bagdatli, S. T., Choudhry, H., Chang, H. Y., & Greenleaf, W. J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature Genetics*, *53*(3), 403–411. https://doi.org/10.1038/s41588-021-00790-6

Granja, J. M., Klemm, S., McGinnis, L. M., Kathiria, A. S., Mezger, A., Corces, M. R., Parks, B., Gars, E., Liedtke, M., Zheng, G. X. Y., Chang, H. Y., Majeti, R., & Greenleaf, W. J. (2019). Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nature biotechnology*, *37*(12), 1458–1465. https://doi.org/10.1038/s41587-019-0332-7

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics*, *27*(7), 1017–1018. https://doi.org/10.1093/bioinformatics/btr064

Greaves, M. (2005). In utero origins of childhood leukaemia. *Early Human Development*, *81*(1), 123–129. https://doi.org/10.1016/j.earlhumdev.2004.10.004

Greaves, M. (2018). A causal mechanism for childhood acute lymphoblastic leukaemia. *Nature Reviews. Cancer*, *18*(8), 471–484. https://doi.org/10.1038/s41568-018-0015-6

Greil, J., Gramatzki, M., Burger, R., Marschalek, R., Peltner, M., Trautmann, U., Hansen-Hagge, T. E., Bartram, C. R., Fey, G. H., & Stehr, K. (1994). The acute lymphoblastic leukaemia cell line SEM with t(4;11) chromosomal rearrangement is biphenotypic and responsive to interleukin-7. *British Journal of Haematology*, *86*(2), 275–283. https://doi.org/10.1111/j.1365-2141.1994.tb04726.x

Gross, D. S., & Garrard, W. T. (1988). Nuclease hypersensitive sites in chromatin. *Annual Review of Biochemistry*, *57*, 159–197. https://doi.org/10.1146/annurev.bi.57.070188.001111

Growney, J. D., Shigematsu, H., Li, Z., Lee, B. H., Adelsperger, J., Rowan, R., Curley, D. P., Kutok, J. L., Akashi, K., Williams, I. R., Speck, N. A., & Gilliland, D. G. (2005). Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood*, *106*(2), 494–504. https://doi.org/10.1182/blood-2004-08-3280

Guenther, M. G., Lawton, L. N., Rozovskaia, T., Frampton, G. M., Levine, S. S., Volkert, T. L., Croce, C. M., Nakamura, T., Canaani, E., & Young, R. A. (2008). Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. *Genes & Development*, *22*(24), 3403–3408. https://doi.org/10.1101/gad.1741408

Guiu, J., Shimizu, R., D'Altri, T., Fraser, S. T., Hatakeyama, J., Bresnick, E. H., Kageyama, R., Dzierzak, E., Yamamoto, M., Espinosa, L., & Bigas, A. (2013). Hes repressors are essential regulators of hematopoietic stem cell development downstream of Notch signaling. *The Journal of Experimental Medicine*, *210*(1), 71–84. https://doi.org/10.1084/jem.20120993

Gustafsson, C., De Paepe, A., Schmidl, C., & Månsson, R. (2019). High-throughput ChIPmentation: Freely scalable, single day ChIPseq data generation from very low cell-numbers. *BMC Genomics*, *20*(1), 59. https://doi.org/10.1186/s12864-018-5299-0

Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology*, *19*(10), 621–637. https://doi.org/10.1038/s41580-018-0028-8

Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology*, *20*(1), 296. https://doi.org/10.1186/s13059-019-1874-1

Hahn, M. W., & Kern, A. D. (2005). Comparative Genomics of Centrality and Essentiality in Three Eukaryotic Protein-Interaction Networks. *Molecular Biology and Evolution*, *22*(4), 803–806. https://doi.org/10.1093/molbev/msi072

Halfon, M. S., Carmena, A., Gisselbrecht, S., Sackerson, C. M., Jiménez, F., Baylies, M. K., & Michelson, A. M. (2000). Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell*, *103*(1), 63–74. https://doi.org/10.1016/s0092-8674(00)00105-7

Hanna-Rose, W., & Hansen, U. (1996). Active repression mechanisms of eukaryotic transcription repressors. *Trends in genetics: TIG*, *12*(6), 229–234. https://doi.org/10.1016/0168-9525(96)10022-6

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., ... Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

Hariprakash, J. M., & Ferrari, F. (2019). Computational Biology Solutions to Identify Enhancers-target Gene Pairs. *Computational and Structural Biotechnology Journal*, *17*, 821–831. https://doi.org/10.1016/j.csbj.2019.06.012

Harland, L. T. G., Simon, C. S., Senft, A. D., Costello, I., Greder, L., Imaz-Rosshandler, I., Göttgens, B., Marioni, J. C., Bikoff, E. K., Porcher, C., de Bruijn, M. F. T. R., & Robertson, E. J. (2021). The T-box transcription factor Eomesodermin governs haemogenic competence of yolk sac mesodermal progenitors. *Nature Cell Biology*, *23*(1), 61–74. https://doi.org/10.1038/s41556-020-00611-8

Harman, J. R., Thorne, R., Jamilly, M., Tapia, M., Crump, N. T., Rice, S., Beveridge, R., Morrissey, E., de Bruijn, M. F. T. R., Roberts, I., Roy, A., Fulga, T. A., & Milne, T. A. (2021). A KMT2A-AFF1 gene regulatory network highlights the role of core transcription factors and reveals the regulatory logic of key downstream target genes. *Genome Research*. https://doi.org/10.1101/gr.268490.120

Hashimshony, T., Zhang, J., Keshet, I., Bustin, M., & Cedar, H. (2003). The role of DNA methylation in setting up chromatin structure during development. *Nature Genetics*, *34*(2), 187–192. https://doi.org/10.1038/ng1158

Hass, M. R., Brissette, D., Parameswaran, S., Pujato, M., Donmez, O., Kottyan, L. C., Weirauch, M. T., & Kopan, R. (2021). Runx1 shapes the chromatin landscape via a cascade of direct and indirect targets. *PLoS Genetics*, *17*(6), e1009574. https://doi.org/10.1371/journal.pgen.1009574

Heaton, H., Talman, A. M., Knights, A., Imaz, M., Gaffney, D. J., Durbin, R., Hemberg, M., & Lawniczak, M. K. N. (2020). Souporcell: Robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nature Methods*, *17*(6), 615–620. https://doi.org/10.1038/s41592-020-0820-1

Hecker, M., Lambeck, S., Toepfer, S., van Someren, E., & Guthke, R. (2009). Gene regulatory network inference: Data integration in dynamic models—A review. *Biosystems*, *96*(1), 86–103. https://doi.org/10.1016/j.biosystems.2008.12.004

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz, J. E., Liu, H., Zhang, X., Green, R. D., Stewart, R., Thomson, J. A., Crawford, G. E., Kellis, M., & Ren, B. (2009). Histone Modifications at Human Enhancers Reflect

Global Cell Type-Specific Gene Expression. *Nature*, *459*(7243), 108–112. https://doi.org/10.1038/nature07829

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., & Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, *39*(3), 311–318. https://doi.org/10.1038/ng1966

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, *38*(4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Hilden, J. M., Dinndorf, P. A., Meerbaum, S. O., Sather, H., Villaluna, D., Heerema, N. A., McGlennen, R., Smith, F. O., Woods, W. G., Salzer, W. L., Johnstone, H. S., Dreyer, Z., Reaman, G. H., & Children's Oncology Group. (2006). Analysis of prognostic factors of acute lymphoblastic leukemia in infants: Report on CCG 1953 from the Children's Oncology Group. *Blood*, *108*(2), 441–451. https://doi.org/10.1182/blood-2005-07-3011

Hisa, T., Spence, S. E., Rachel, R. A., Fujita, M., Nakamura, T., Ward, J. M., Devor-Henneman, D. E., Saiki, Y., Kutsuna, H., Tessarollo, L., Jenkins, N. A., & Copeland, N. G. (2004). Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. *The EMBO Journal*, *23*(2), 450–459. https://doi.org/10.1038/sj.emboj.7600038

Hope, I. A., & Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell*, *46*(6), 885–894. https://doi.org/10.1016/0092-8674(86)90070-x

Hosokawa, H., Koizumi, M., Masuhara, K., Romero-Wolf, M., Tanaka, T., Nakayama, T., & Rothenberg, E. V. (2021). Stage-specific action of Runx1 and GATA3 controls silencing of PU.1 expression in mouse pro–T cells. *Journal of Experimental Medicine*, *218*(8), e20202648. https://doi.org/10.1084/jem.20202648

Hou, S., Li, Z., Zheng, X., Gao, Y., Dong, J., Ni, Y., Wang, X., Li, Y., Ding, X., Chang, Z., Li, S., Hu, Y., Fan, X., Hou, Y., Wen, L., Liu, B., Tang, F., & Lan, Y. (2020). Embryonic endothelial evolution towards first hematopoietic stem cells revealed by single-cell transcriptomic and functional analyses. *Cell Research*, *30*(5), 376–392. https://doi.org/10.1038/s41422-020-0300-2

Hu, Z., Gu, X., Baraoidan, K., Ibanez, V., Sharma, A., Kadkol, S., Munker, R., Ackerman, S., Nucifora, G., & Sauntharajah, Y. (2011). RUNX1 regulates

corepressor interactions of PU.1. *Blood*, *117*(24), 6498–6508. https://doi.org/10.1182/blood-2010-10-312512

Huang, D., Petrykowska, H. M., Miller, B. F., Elnitski, L., & Ovcharenko, I. (2019). Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. *Genome Research*, *29*(4), 657–667. https://doi.org/10.1101/gr.247007.118

Huang, G., Zhang, P., Hirai, H., Elf, S., Yan, X., Chen, Z., Koschmieder, S., Okuno, Y., Dayaram, T., Growney, J. D., Shivdasani, R. A., Gilliland, D. G., Speck, N. A., Nimer, S. D., & Tenen, D. G. (2008). PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. *Nature Genetics*, *40*(1), 51–60. https://doi.org/10.1038/ng.2007.7

Huang, K., Gao, J., Du, J., Ma, N., Zhu, Y., Wu, P., Zhang, T., Wang, W., Li, Y., Chen, Q., Hutchins, A. P., Yang, Z., Zheng, Y., Zhang, J., Shan, Y., Li, X., Liao, B., Liu, J., Wang, J., … Pan, G. (2016). Generation and Analysis of GATA2w/eGFP Human ESCs Reveal ITGB3/CD61 as a Reliable Marker for Defining Hemogenic Endothelial Cells during Hematopoiesis. *Stem Cell Reports*, *7*(5), 854–868. https://doi.org/10.1016/j.stemcr.2016.09.008

Huang, Z., Chen, K., Chi, Y., Jin, H., Li, L., Zhang, W., Xu, J., & Zhang, Y. (2021). Runx1 regulates zebrafish neutrophil maturation via synergistic interaction with c-Myb. *The Journal of Biological Chemistry*, *296*, 100272. https://doi.org/10.1016/j.jbc.2021.100272

Hughes, C. M., Rozenblatt-Rosen, O., Milne, T. A., Copeland, T. D., Levine, S. S., Lee, J. C., Hayes, D. N., Shanmugam, K. S., Bhattacharjee, A., Biondi, C. A., Kay, G. F., Hayward, N. K., Hess, J. L., & Meyerson, M. (2004). Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. *Molecular Cell*, *13*(4), 587–597. https://doi.org/10.1016/s1097-2765(04)00081-4

Huntzinger, E., & Izaurralde, E. (2011). Gene silencing by microRNAs: Contributions of translational repression and mRNA decay. *Nature Reviews Genetics*, *12*(2), 99–110. https://doi.org/10.1038/nrg2936

Iacovino, M., Chong, D., Szatmari, I., Hartweck, L., Rux, D., Caprioli, A., Cleaver, O., & Kyba, M. (2011). HoxA3 is an apical regulator of haemogenic endothelium. *Nature Cell Biology*, *13*(1), 72–78. https://doi.org/10.1038/ncb2137

Ichikawa, M., Asai, T., Saito, T., Seo, S., Yamazaki, I., Yamagata, T., Mitani, K., Chiba, S., Ogawa, S., Kurokawa, M., & Hirai, H. (2004). AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nature Medicine*, *10*(3), 299–304. https://doi.org/10.1038/nm997

Ichikawa, M., Yoshimi, A., Nakagawa, M., Nishimoto, N., Watanabe-Okochi, N., & Kurokawa, M. (2013). A role for RUNX1 in hematopoiesis and myeloid leukemia. *International Journal of Hematology*, *97*(6), 726–734. https://doi.org/10.1007/s12185-013-1347-3

Ideker, T. E., Thorsson, V., & Karp, R. M. (2000). Discovery of regulatory interactions through perturbation: Inference and experimental design. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 305–316. https://doi.org/10.1142/9789814447331_0029

Illendula, A., Gilmour, J., Grembecka, J., Tirumala, V. S. S., Boulton, A., Kuntimaddi, A., Schmidt, C., Wang, L., Pulikkan, J. A., Zong, H., Parlak, M., Kuscu, C., Pickin, A., Zhou, Y., Gao, Y., Mishra, L., Adli, M., Castilla, L. H., Rajewski, R. A., … Bushweller, J. H. (2016). Small Molecule Inhibitor of CBFβ-RUNX Binding for RUNX Transcription Factor Driven Cancers. *EBioMedicine*, *8*, 117–131. https://doi.org/10.1016/j.ebiom.2016.04.032

Illendula, A., Pulikkan, J. A., Zong, H., Grembecka, J., Xue, L., Sen, S., Zhou, Y., Boulton, A., Kuntimaddi, A., Gao, Y., Rajewski, R. A., Guzman, M. L., Castilla, L. H., & Bushweller, J. H. (2015). A small-molecule inhibitor of the aberrant transcription factor CBFβ-SMMHC delays leukemia in mice. *Science (New York, N.Y.)*, *347*(6223), 779–784. https://doi.org/10.1126/science.aaa0314

Imperato, M. R., Cauchy, P., Obier, N., & Bonifer, C. (2015). The RUNX1–PU.1 axis in the control of hematopoiesis. *International Journal of Hematology*, *101*(4), 319–329. https://doi.org/10.1007/s12185-015-1762-8

Ingram, P. J., Stumpf, M. P., & Stark, J. (2006). Network motifs: Structure does not determine function. *BMC Genomics*, *7*, 108. https://doi.org/10.1186/1471-2164-7-108

Jackson, T. R., Ling, R. E., & Roy, A. (2021). The Origin of B-cells: Human Fetal B Cell Development and Implications for the Pathogenesis of Childhood Acute Lymphoblastic Leukemia. *Frontiers in Immunology*, *12*.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLOS ONE*, *9*(6), e98679. https://doi.org/10.1371/journal.pone.0098679

Jang, M. K., Mochizuki, K., Zhou, M., Jeong, H.-S., Brady, J. N., & Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Molecular Cell*, *19*(4), 523–534. https://doi.org/10.1016/j.molcel.2005.06.027

Jardine, L., Webb, S., Goh, I., Quiroga Londoño, M., Reynolds, G., Mather, M., Olabi, B., Stephenson, E., Botting, R. A., Horsfall, D., Engelbert, J., Maunder, D., Mende, N., Murnane, C., Dann, E., McGrath, J., King, H., Kucinski, I., Queen, R., ... Haniffa, M. (2021). Blood and immune development in human fetal bone marrow and Down syndrome. *Nature*, *598*(7880), 327–331. https://doi.org/10.1038/s41586-021-03929-x

Jeong, H., Mason, S. P., Barabási, A.-L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41–42. https://doi.org/10.1038/35075138

Joanito, I., Chu, J.-W., Wu, S.-H., & Hsu, C.-P. (2018). An incoherent feed-forward loop switches the Arabidopsis clock rapidly between two hysteretic states. *Scientific Reports*, *8*(1), 13944. https://doi.org/10.1038/s41598-018-32030-z

Johnson, K. M., Wang, J., Smallwood, A., Arayata, C., & Carey, M. (2002). TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. *Genes & Development*, *16*(14), 1852–1863. https://doi.org/10.1101/gad.995702

Jones, P. A. (2012). Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, *13*(7), 484–492. https://doi.org/10.1038/nrg3230

Joy, M. P., Brock, A., Ingber, D. E., & Huang, S. (2005). High-Betweenness Proteins in the Yeast Protein Interaction Network. *Journal of Biomedicine and Biotechnology*, *2005*(2), 96–103. https://doi.org/10.1155/JBB.2005.96

Kadonaga, J., Courey, A., Ladika, J., & Tjian, R. (1988). Distinct regions of Sp1 modulate DNA binding and transcriptional activation. *Science (New York, N.Y.)*, *242*(4885). https://doi.org/10.1126/science.3059495

Kale, J., Osterlund, E. J., & Andrews, D. W. (2018). BCL-2 family proteins: Changing partners in the dance towards death. *Cell Death & Differentiation*, *25*(1), 65–80. https://doi.org/10.1038/cdd.2017.186

Kalir, S., Mangan, S., & Alon, U. (2005). A coherent feed-forward loop with a SUM input function prolongs flagella expression in Escherichia coli. *Molecular Systems Biology*, *1*, 2005.0006. https://doi.org/10.1038/msb4100010

Karlebach, G., & Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, *9*(10), 770–780. https://doi.org/10.1038/nrm2503

Kataoka, K. (2007). Multiple Mechanisms and Functions of Maf Transcription Factors in the Regulation of Tissue-Specific Genes. *The Journal of Biochemistry*, *141*(6), 775–781. https://doi.org/10.1093/jb/mvm105

Katsuoka, F., & Yamamoto, M. (2016). Small Maf proteins (MafF, MafG, MafK): History, structure and function. *Gene*, *586*(2), 197–205. https://doi.org/10.1016/j.gene.2016.03.058

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, *12*(6), 996–1006. https://doi.org/10.1101/gr.229102

Kerry, J., Godfrey, L., Repapi, E., Tapia, M., Blackledge, N. P., Ma, H., Ballabio, E., O'Byrne, S., Ponthan, F., Heidenreich, O., Roy, A., Roberts, I., Konopleva, M., Klose, R. J., Geng, H., & Milne, T. A. (2017). MLL-AF4 Spreading Identifies Binding Sites that Are Distinct from Super-Enhancers and that Govern Sensitivity to DOT1L Inhibition in Leukemia. *Cell Reports*, *18*(2), 482–495. https://doi.org/10.1016/j.celrep.2016.12.054

Khaw, S. L., Suryani, S., Evans, K., Richmond, J., Robbins, A., Kurmasheva, R. T., Billups, C. A., Erickson, S. W., Guo, Y., Houghton, P. J., Smith, M. A., Carol, H., Roberts, A. W., Huang, D. C. S., & Lock, R. B. (2016). Venetoclax responses of pediatric ALL xenografts reveal sensitivity of MLL-rearranged leukemia. *Blood*, *128*(10), 1382–1395. https://doi.org/10.1182/blood-2016-03-707414

Kim, J., Sif, S., Jones, B., Jackson, A., Koipally, J., Heller, E., Winandy, S., Viel, A., Sawyer, A., Ikeda, T., Kingston, R., & Georgopoulos, K. (1999a). Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. *Immunity*, *10*(3), 345–355. https://doi.org/10.1016/s1074-7613(00)80034-5

Kim, W. Y., Sieweke, M., Ogawa, E., Wee, H. J., Englmeier, U., Graf, T., & Ito, Y. (1999b). Mutual activation of Ets-1 and AML1 DNA binding by direct interaction of their autoinhibitory domains. *The EMBO Journal*, *18*(6), 1609–1620. https://doi.org/10.1093/emboj/18.6.1609

Kissa, K., & Herbomel, P. (2010). Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature*, *464*(7285), 112–115. https://doi.org/10.1038/nature08761

Kitabayashi, I., Yokoyama, A., Shimizu, K., & Ohki, M. (1998). Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. *The EMBO Journal*, *17*(11), 2994. https://doi.org/10.1093/emboj/17.11.2994

Kleinmann, E., Geimer Le Lay, A.-S., Sellars, M., Kastner, P., & Chan, S. (2008). Ikaros Represses the Transcriptional Response to Notch Signaling in T-Cell Development. *Molecular and Cellular Biology*, *28*(24), 7465–7475. https://doi.org/10.1128/MCB.00715-08

Klemm, S. L., Shipony, Z., & Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics*, *20*(4), 207–220. https://doi.org/10.1038/s41576-018-0089-8

Kojika, S., & Griffin, J. D. (2001). Notch receptors and hematopoiesis. *Experimental Hematology*, *29*(9), 1041–1052. https://doi.org/10.1016/s0301-472x(01)00676-2

Komatsu, M., Li, H.-O., Tsutsui, H., Itakura, K., Matsumura, M., & Yokoyama, K. K. (1997). MAZ, a Myc-associated zinc finger protein, is essential for the ME1a1-mediated expression of the c-myc gene during neuroectodermal differentiation of P19 cells. *Oncogene*, *15*(10), 1123–1131. https://doi.org/10.1038/sj.onc.1201283

Koschützki, D., & Schreiber, F. (2008). Centrality Analysis Methods for Biological Networks and Their Application to Gene Regulatory Networks. *Gene Regulation and Systems Biology*, *2*, 193–201.

Krackhardt, D. (1990). Assessing the Political Landscape: Structure, Cognition, and Power in Organizations. *Administrative Science Quarterly*, *35*(2), 342–369. https://doi.org/10.2307/2393394

Krivtsov, A. V., Feng, Z., Lemieux, M. E., Faber, J., Vempati, S., Sinha, A. U., Xia, X., Jesneck, J., Bracken, A. P., Silverman, L. B., Kutok, J. L., Kung, A. L., & Armstrong, S. A. (2008). H3K79 methylation profiles define murine and human MLL-AF4 leukemias. *Cancer Cell*, *14*(5), 355–368. https://doi.org/10.1016/j.ccr.2008.10.001

Krivtsov, A. V., Figueroa, M. E., Sinha, A. U., Stubbs, M. C., Feng, Z., Valk, P. J., Delwel, R., Döhner, K., Bullinger, L., Kung, A. L., Melnick, A. M., & Armstrong, S. A. (2013). Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. *Leukemia*, *27*(4), 852–860. https://doi.org/10.1038/leu.2012.363

Kueh, H. Y., & Rothenberg, E. V. (2012). Regulatory gene network circuits underlying T cell development from multipotent progenitors. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine*, *4*(1), 79–102. https://doi.org/10.1002/wsbm.162

Kulakovskiy, I. V., Vorontsov, I. E., Yevshin, I. S., Sharipov, R. N., Fedorova, A. D., Rumynskiy, E. I., Medvedeva, Y. A., Magana-Mora, A., Bajic, V. B., Papatsenko, D. A., Kolpakov, F. A., & Makeev, V. J. (2018). HOCOMOCO: Towards a complete collection of transcription factor binding models for human

and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Research*, *46*(D1), D252–D259. https://doi.org/10.1093/nar/gkx1106

Kuvardina, O. N., Herglotz, J., Kolodziej, S., Kohrs, N., Herkt, S., Wojcik, B., Oellerich, T., Corso, J., Behrens, K., Kumar, A., Hussong, H., Urlaub, H., Koch, J., Serve, H., Bonig, H., Stocking, C., Rieger, M. A., & Lausen, J. (2015). RUNX1 represses the erythroid gene expression program during megakaryocytic differentiation. *Blood*, *125*(23), 3570–3579. https://doi.org/10.1182/blood-2014-11-610519

Lacoste, N., Utley, R. T., Hunter, J. M., Poirier, G. G., & Côte, J. (2002). Disruptor of telomeric silencing-1 is a chromatin-specific histone H3 methyltransferase. *The Journal of Biological Chemistry*, *277*(34), 30421–30424. https://doi.org/10.1074/jbc.C200366200

Lancrin, C., Mazan, M., Stefanska, M., Patel, R., Lichtinger, M., Costa, G., Vargel, Ö., Wilson, N. K., Möröy, T., Bonifer, C., Göttgens, B., Kouskoff, V., & Lacaud, G. (2012). GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood*, *120*(2), 314–322. https://doi.org/10.1182/blood-2011-10-386094

Landherr, A., Friedl, B., & Heidemann, J. (2010). A Critical Review of Centrality Measures in Social Networks. *Business & Information Systems Engineering*, *2*(6), 371–385. https://doi.org/10.1007/s12599-010-0127-3

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559. https://doi.org/10.1186/1471-2105-9-559

Langmead, B. (2010). Aligning short sequencing reads with Bowtie. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]*, *CHAPTER*, Unit–11.7. https://doi.org/10.1002/0471250953.bi1107s32

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Latchman, D. S. (1997). Transcription factors: An overview. *The International Journal of Biochemistry & Cell Biology*, *29*(12), 1305–1312. https://doi.org/10.1016/S1357-2725(97)00085-X

Lau, I. J. (2022). *The role of MYB in MLL-AF9 leukaemia* (http://purl.org/dc/dcmitype/Text). University of Oxford.

Lebrin, F., Deckers, M., Bertolino, P., & ten Dijke, P. (2005). TGF-β receptor function in the endothelium. *Cardiovascular Research*, *65*(3), 599–608. https://doi.org/10.1016/j.cardiores.2004.10.036

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. J., Tagne, J.-B., Volkert, T. L., Fraenkel, E., … Young, R. A. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science (New York, N.Y.)*, *298*(5594), 799–804. https://doi.org/10.1126/science.1075090

Lee, T. I., & Young, R. A. (2013). Transcriptional Regulation and its Misregulation in Disease. *Cell*, *152*(6), 1237–1251. https://doi.org/10.1016/j.cell.2013.02.014

Lehner, B., Crombie, C., Tischler, J., Fortunato, A., & Fraser, A. G. (2006). Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. *Nature Genetics*, *38*(8), 896–903. https://doi.org/10.1038/ng1844

Lehnertz, B., Chagraoui, J., MacRae, T., Tomellini, E., Corneau, S., Mayotte, N., Boivin, I., Durand, A., Gracias, D., & Sauvageau, G. (2021). HLF expression defines the human hematopoietic stem cell state. *Blood*, *138*(25), 2642–2654. https://doi.org/10.1182/blood.2021010745

Levine, M., & Davidson, E. H. (2005). Gene regulatory networks for development. *Proceedings of the National Academy of Sciences*, *102*(14), 4936–4942. https://doi.org/10.1073/pnas.0408031102

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J., Riedt, T., Goossens, S., Carrillo García, C., Szczepanski, S., Brandes, M., Pieters, T., Dobrosch, L., Gütgemann, I., Farla, N., Radaelli, E., Hulpiau, P., Mallela, N., Fröhlich, H., La Starza, R., Matteucci, C., Chen, T., Brossart, P., Mecucci, C., … Janzen, V. (2017). The EMT transcription factor Zeb2 controls adult murine hematopoietic differentiation by regulating cytokine signaling. *Blood*, *129*(4), 460–472. https://doi.org/10.1182/blood-2016-05-714659

Liang, P., Hu, R., Liu, Z., Miao, M., Jiang, H., & Li, C. (2020). NAT10 upregulation indicates a poor prognosis in acute myeloid leukemia. *Current Problems in Cancer*, *44*(2), 100491. https://doi.org/10.1016/j.currproblcancer.2019.06.006

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Licht, J. D., Grossel, M. J., Figge, J., & Hansen, U. M. (1990). Drosophila Krüppel protein is a transcriptional represser. *Nature*, *346*(6279), 76–79. https://doi.org/10.1038/346076a0

Lichtinger, M., Ingram, R., Hannah, R., Müller, D., Clarke, D., Assi, S. A., Lie-A-Ling, M., Noailles, L., Vijayabaskar, M. S., Wu, M., Tenen, D. G., Westhead, D. R., Kouskoff, V., Lacaud, G., Göttgens, B., & Bonifer, C. (2012). RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *The EMBO journal*, *31*(22), 4318–4333. https://doi.org/10.1038/emboj.2012.275

Lie-A-Ling, M., Marinopoulou, E., Li, Y., Patel, R., Stefanska, M., Bonifer, C., Miller, C., Kouskoff, V., & Lacaud, G. (2014). RUNX1 positively regulates a cell adhesion and migration program in murine hemogenic endothelium prior to blood emergence. *Blood*, *124*(11), e11–20. https://doi.org/10.1182/blood-2014-04-572958

Lie-A-Ling, M., Marinopoulou, E., Lilly, A. J., Challinor, M., Patel, R., Lancrin, C., Kouskoff, V., & Lacaud, G. (2018). Regulation of RUNX1 dosage is crucial for efficient blood formation from hemogenic endothelium. *Development (Cambridge, England)*, *145*(5), dev149419. https://doi.org/10.1242/dev.149419

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, *326*(5950), 289–293. https://doi.org/10.1126/science.1181369

Lillie, J. W., Green, M., & Green, M. R. (1986). An adenovirus E1a protein region required for transformation and transcriptional repression. *Cell*, *46*(7), 1043–1051. https://doi.org/10.1016/0092-8674(86)90704-x

Lin, C., Garrett, A. S., De Kumar, B., Smith, E. R., Gogol, M., Seidel, C., Krumlauf, R., & Shilatifard, A. (2011). Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). *Genes & Development*, *25*(14), 1486–1498. https://doi.org/10.1101/gad.2059211

Lin, C., Smith, E. R., Takahashi, H., Lai, K. C., Martin-Brown, S., Florens, L., Washburn, M. P., Conaway, J. W., Conaway, R. C., & Shilatifard, A. (2010). AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit

of MLL chimeras, can link transcription elongation to leukemia. *Molecular Cell*, *37*(3), 429–437. https://doi.org/10.1016/j.molcel.2010.01.026

Lin, S., Luo, R. T., Ptasinska, A., Kerry, J., Assi, S. A., Wunderlich, M., Imamura, T., Kaberlein, J. J., Rayes, A., Althoff, M. J., Anastasi, J., O'Brien, M. M., Meetei, A. R., Milne, T. A., Bonifer, C., Mulloy, J. C., & Thirman, M. J. (2016). Instructive Role of MLL-Fusion Proteins Revealed by a Model of t(4;11) Pro-B Acute Lymphoblastic Leukemia. *Cancer Cell*, *30*(5), 737–749. https://doi.org/10.1016/j.ccell.2016.10.008

Ling, K.-W., Ottersbach, K., van Hamburg, J. P., Oziemlak, A., Tsai, F.-Y., Orkin, S. H., Ploemacher, R., Hendriks, R. W., & Dzierzak, E. (2004). GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *The Journal of Experimental Medicine*, *200*(7), 871–882. https://doi.org/10.1084/jem.20031556

Liu, F., Li, D., Yu, Y. Y. L., Kang, I., Cha, M.-J., Kim, J. Y., Park, C., Watson, D. K., Wang, T., & Choi, K. (2015). Induction of hematopoietic and endothelial cell program orchestrated by ETS transcription factor ER71/ETV2. *EMBO reports*, *16*(5), 654–669. https://doi.org/10.15252/embr.201439939

Liu, Z., Vong, Q. P., Liu, C., & Zheng, Y. (2014). Borg5 is required for angiogenesis by regulating persistent directional migration of the cardiac microvascular endothelial cells. *Molecular Biology of the Cell*, *25*(6), 841–851. https://doi.org/10.1091/mbc.E13-09-0543

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)*, *25*(4), 402–408. https://doi.org/10.1006/meth.2001.1262

Lizama, C. O., Hawkins, J. S., Schmitt, C. E., Bos, F. L., Zape, J. P., Cautivo, K. M., Borges Pinto, H., Rhyner, A. M., Yu, H., Donohoe, M. E., Wythe, J. D., & Zovein, A. C. (2015). Repression of arterial genes in hemogenic endothelium is sufficient for haematopoietic fate acquisition. *Nature Communications*, *6*, 7739. https://doi.org/10.1038/ncomms8739

Lock, L. F., Takagi, N., & Martin, G. R. (1987). Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell*, *48*(1), 39–46. https://doi.org/10.1016/0092-8674(87)90353-9

Lopes-Ramos, C. M., Paulson, J. N., Chen, C.-Y., Kuijjer, M. L., Fagny, M., Platig, J., Sonawane, A. R., DeMeo, D. L., Quackenbush, J., & Glass, K. (2017). Regulatory network changes between cell lines and their tissues of origin. *BMC Genomics*, *18*(1), 723. https://doi.org/10.1186/s12864-017-4111-x

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, S.-J., Li, F., Yin, H., Feng, Q., Kimbrel, E. A., Hahm, E., Thon, J. N., Wang, W., Italiano, J. E., Cho, J., & Lanza, R. (2011). Platelets generated from human embryonic stem cells are functional in vitro and in the microcirculation of living mice. *Cell Research*, *21*(3), 530–545. https://doi.org/10.1038/cr.2011.8

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Luger, K., & Richmond, T. J. (1998). The histone tails of the nucleosome. *Current Opinion in Genetics & Development*, *8*(2), 140–146. https://doi.org/10.1016/s0959-437x(98)80134-2

Luscombe, N. M., Austin, S. E., Berman, H. M., & Thornton, J. M. (2000). An overview of the structures of protein-DNA complexes. *Genome Biology*, *1*(1), reviews001.1–reviews001.37.

Lutterbach, B., & Hiebert, S. W. (2000). Role of the transcription factor AML-1 in acute leukemia and hematopoietic differentiation. *Gene*, *245*(2), 223–235. https://doi.org/10.1016/S0378-1119(00)00014-7

Ma, J., & Ptashne, M. (1987). A new class of yeast transcriptional activators. *Cell*, *51*(1), 113–119. https://doi.org/10.1016/0092-8674(87)90015-8

Machanick, P., & Bailey, T. L. (2011). MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics (Oxford, England)*, *27*(12), 1696–1697. https://doi.org/10.1093/bioinformatics/btr189

MacNeil, L. T., & Walhout, A. J. (2011). Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Research*, *21*(5), 645–657. https://doi.org/10.1101/gr.097378.109

Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics (Oxford, England)*, *27*(21), 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

Mak, A. B., Nixon, A. M. L., & Moffat, J. (2012). The Mixed Lineage Leukemia (MLL) Fusion–Associated Gene AF4 Promotes CD133 Transcription. *Cancer Research*, *72*(8), 1929–1934. https://doi.org/10.1158/0008-5472.CAN-11-3589

Mak, K. S., Funnell, A. P. W., Pearson, R. C. M., & Crossley, M. (2011). PU.1 and Haematopoietic Cell Fate: Dosage Matters. *International Journal of Cell Biology*, *2011*, 808524. https://doi.org/10.1155/2011/808524

Mangan, S., & Alon, U. (2003a). Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences*, *100*(21), 11980–11985. https://doi.org/10.1073/pnas.2133841100

Mangan, S., Itzkovitz, S., Zaslaver, A., & Alon, U. (2006). The Incoherent Feed-forward Loop Accelerates the Response-time of the gal System of Escherichia coli. *Journal of Molecular Biology*, *356*(5), 1073–1081. https://doi.org/10.1016/j.jmb.2005.12.003

Mangan, S., Zaslaver, A., & Alon, U. (2003b). The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *Journal of Molecular Biology*, *334*(2), 197–204. https://doi.org/10.1016/j.jmb.2003.09.049

Mann, G., Attarbaschi, A., Schrappe, M., De Lorenzo, P., Peters, C., Hann, I., De Rossi, G., Felice, M., Lausen, B., Leblanc, T., Szczepanski, T., Ferster, A., Janka-Schaub, G., Rubnitz, J., Silverman, L. B., Stary, J., Campbell, M., Li, C. K., Suppiah, R., … Interfant-99 Study Group. (2010). Improved outcome with hematopoietic stem cell transplantation in a poor prognostic subgroup of infants with mixed-lineage-leukemia (MLL)-rearranged acute lymphoblastic leukemia: Results from the Interfant-99 Study. *Blood*, *116*(15), 2644–2650. https://doi.org/10.1182/blood-2010-03-273532

Marke, R., van Leeuwen, F. N., & Scheijen, B. (2018). The many faces of IKZF1 in B-cell precursor acute lymphoblastic leukemia. *Haematologica*, *103*(4), 565–574. https://doi.org/10.3324/haematol.2017.185603

Marsman, J., Thomas, A., Osato, M., O'Sullivan, J. M., & Horsfield, J. A. (2017). A DNA Contact Map for the Mouse Runx1 Gene Identifies Novel Haematopoietic Enhancers. *Scientific Reports*, *7*(1), 13347. https://doi.org/10.1038/s41598-017-13748-8

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads [Number: 1]. *EMBnet.journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Martinez, M., Hinojosa, M., Trombly, D., Morin, V., Stein, J., Stein, G., Javed, A., & Gutierrez, S. E. (2016). Transcriptional Auto-Regulation of RUNX1 P1 Promoter. *PLoS ONE*, *11*(2), e0149119. https://doi.org/10.1371/journal.pone.0149119

Materna, S. C., & Davidson, E. H. (2007). Logic of gene regulatory networks. *Current opinion in biotechnology*, *18*(4), 351–354. https://doi.org/10.1016/j.copbio.2007.07.008

Mayall, T. P., Sheridan, P. L., Montminy, M. R., & Jones, K. A. (1997). Distinct roles for P-CREB and LEF-1 in TCR alpha enhancer assembly and activation on chromatin templates in vitro. *Genes & Development*, *11*(7), 887–899. https://doi.org/10.1101/gad.11.7.887

McCarthy, D. J., Campbell, K. R., Lun, A. T. L., & Wills, Q. F. (2017). Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics*, *33*(8), 1179–1186. https://doi.org/10.1093/bioinformatics/btw777

McGrath, K. E., Frame, J. M., Fegan, K. H., Bowen, J. R., Conway, S. J., Catherman, S. C., Kingsley, P. D., Koniski, A. D., & Palis, J. (2015). Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo. *Cell Reports*, *11*(12), 1892–1904. https://doi.org/10.1016/j.celrep.2015.05.036

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M., & Bejerano, G. (2010). GREAT improves functional interpretation of cis -regulatory regions. *Nature Biotechnology*, *28*(5), 495–501. https://doi.org/10.1038/nbt.1630

McLeay, R. C., & Bailey, T. L. (2010). Motif Enrichment Analysis: A unified framework and an evaluation on ChIP data. *BMC Bioinformatics*, *11*(1), 165. https://doi.org/10.1186/1471-2105-11-165

McMahon, K. A., Hiew, S. Y.-L., Hadjur, S., Veiga-Fernandes, H., Menzel, U., Price, A. J., Kioussis, D., Williams, O., & Brady, H. J. M. (2007). Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal. *Cell Stem Cell*, *1*(3), 338–345. https://doi.org/10.1016/j.stem.2007.07.002

Medvedovic, J., Ebert, A., Tagoh, H., & Busslinger, M. (2011). Pax5: A master regulator of B cell development and leukemogenesis. *Advances in Immunology*, *111*, 179–206. https://doi.org/10.1016/B978-0-12-385991-4.00005-2

Medvinsky, A., & Dzierzak, E. (1996). Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell*, *86*(6), 897–906. https://doi.org/10.1016/s0092-8674(00)80165-8

Menegatti, S., de Kruijf, M., Garcia-Alegria, E., Lacaud, G., & Kouskoff, V. (2019). Transcriptional control of blood cell emergence [_eprint:

https://onlinelibrary.wiley.com/doi/pdf/10.1002/1873-3468.13585]. *FEBS Letters*, *593*(23), 3304–3315. https://doi.org/10.1002/1873-3468.13585

Mercatelli, D., Scalambra, L., Triboli, L., Ray, F., & Giorgi, F. M. (2020). Gene regulatory network inference resources: A practical overview. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1863*(6), 194430. https://doi.org/10.1016/j.bbagrm.2019.194430

Meulemans, D., & Bronner-Fraser, M. (2004). Gene-regulatory interactions in neural crest evolution and development. *Developmental Cell*, *7*(3), 291–299. https://doi.org/10.1016/j.devcel.2004.08.007

Meyer, C., Burmeister, T., Gröger, D., Tsaur, G., Fechina, L., Renneville, A., Sutton, R., Venn, N. C., Emerenciano, M., Pombo-de-Oliveira, M. S., Barbieri Blunck, C., Almeida Lopes, B., Zuna, J., Trka, J., Ballerini, P., Lapillonne, H., De Braekeleer, M., Cazzaniga, G., Corral Abascal, L., ... Marschalek, R. (2018). The MLL recombinome of acute leukemias in 2017. *Leukemia*, *32*(2), 273–284. https://doi.org/10.1038/leu.2017.213

Meyer, C., Hofmann, J., Burmeister, T., Gröger, D., Park, T. S., Emerenciano, M., Pombo de Oliveira, M., Renneville, A., Villarese, P., Macintyre, E., Cavé, H., Clappier, E., Mass-Malo, K., Zuna, J., Trka, J., De Braekeleer, E., De Braekeleer, M., Oh, S. H., Tsaur, G., ... Marschalek, R. (2013). The MLL recombinome of acute leukemias in 2013. *Leukemia*, *27*(11), 2165–2176. https://doi.org/10.1038/leu.2013.135

Meyers, R. M., Bryan, J. G., McFarland, J. M., Weir, B. A., Sizemore, A. E., Xu, H., Dharia, N. V., Montgomery, P. G., Cowley, G. S., Pantel, S., Goodale, A., Lee, Y., Ali, L. D., Jiang, G., Lubonja, R., Harrington, W. F., Strickland, M., Wu, T., Hawes, D. C., ... Tsherniak, A. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nature Genetics*, *49*(12), 1779–1784. https://doi.org/10.1038/ng.3984

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S. W., Andrews, S., Grey, W., Ewels, P. A., Herman, B., Happe, S., Higgs, A., LeProust, E., Follows, G. A., Fraser, P., Luscombe, N. M., & Osborne, C. S. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics*, *47*(6), 598–606. https://doi.org/10.1038/ng.3286

Mii, S., Enomoto, A., Shiraki, Y., Taki, T., Murakumo, Y., & Takahashi, M. (2019). CD109: A multifunctional GPI-anchored protein with key roles in tumor progression and physiological homeostasis. *Pathology International*, *69*(5), 249–259. https://doi.org/10.1111/pin.12798

Millauer, B., Wizigmann-Voos, S., Schnürch, H., Martinez, R., Møller, N. P. H., Risau, W., & Ullrich, A. (1993). High affinity VEGF binding and developmental expression suggest Flk-1 as a major regulator of vasculogenesis and angiogenesis. *Cell*, *72*(6), 835–846. https://doi.org/10.1016/0092-8674(93)90573-9

Miller, M. (2009). The Importance of Being Flexible: The Case of Basic Region Leucine Zipper Transcriptional Regulators. *Current protein & peptide science*, *10*(3), 244–269.

Milne, T. A. (2017). Mouse models of MLL leukemia: Recapitulating the human disease. *Blood*, *129*(16), 2217–2223. https://doi.org/10.1182/blood-2016-10-691428

Milne, T. A., Briggs, S. D., Brock, H. W., Martin, M. E., Gibbs, D., Allis, C. D., & Hess, J. L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. *Molecular Cell*, *10*(5), 1107–1117. https://doi.org/10.1016/s1097-2765(02)00741-4

Milne, T. A., Hughes, C. M., Lloyd, R., Yang, Z., Rozenblatt-Rosen, O., Dou, Y., Schnepp, R. W., Krankel, C., Livolsi, V. A., Gibbs, D., Hua, X., Roeder, R. G., Meyerson, M., & Hess, J. L. (2005a). Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(3), 749–754. https://doi.org/10.1073/pnas.0408836102

Milne, T. A., Kim, J., Wang, G. G., Stadler, S. C., Basrur, V., Whitcomb, S. J., Wang, Z., Ruthenburg, A. J., Elenitoba-Johnson, K. S. J., Roeder, R. G., & Allis, C. D. (2010). Multiple Interactions Recruit MLL1 and MLL1 Fusion Proteins to the HOXA9 Locus in Leukemogenesis. *Molecular Cell*, *38*(6), 853–863. https://doi.org/10.1016/j.molcel.2010.05.011

Milne, T. A., Martin, M. E., Brock, H. W., Slany, R. K., & Hess, J. L. (2005b). Leukemogenic MLL fusion proteins bind across a broad region of the Hox a9 locus, promoting transcription and multiple histone modifications. *Cancer Research*, *65*(24), 11367–11374. https://doi.org/10.1158/0008-5472.CAN-05-1041

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science*, *298*(5594), 824–827. https://doi.org/10.1126/science.298.5594.824

Minehata, K.-I., Kawahara, A., & Suzuki, T. (2008). Meis1 regulates the development of endothelial cells in zebrafish. *Biochemical and Biophysical Research Communications*, *374*(4), 647–652. https://doi.org/10.1016/j.bbrc.2008.07.075

Mirshekar-Syahkal, B., Haak, E., Kimber, G. M., van Leusden, K., Harvey, K., O'Rourke, J., Laborda, J., Bauer, S. R., de Bruijn, M. F. T. R., Ferguson-Smith, A. C., Dzierzak, E., & Ottersbach, K. (2013). Dlk1 is a negative regulator of emerging hematopoietic stem and progenitor cells. *Haematologica*, *98*(2), 163–171. https://doi.org/10.3324/haematol.2012.070789

Mitchell, P. J., & Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science (New York, N.Y.)*, *245*(4916), 371–378. https://doi.org/10.1126/science.2667136

Monteiro, R., Pinheiro, P., Joseph, N., Peterkin, T., Koth, J., Repapi, E., Bonkhofer, F., Kirmizitas, A., & Patient, R. (2016). Transforming Growth Factor β Drives Hemogenic Endothelium Programming and the Transition to Hematopoietic Stem Cells. *Developmental Cell*, *38*(4), 358–370. https://doi.org/10.1016/j.devcel.2016.06.024

Morgan, D., Studham, M., Tjärnberg, A., Weishaupt, H., Swartling, F. J., Nordling, T. E. M., & Sonnhammer, E. L. L. (2020). Perturbation-based gene regulatory network inference to unravel oncogenic mechanisms. *Scientific Reports*, *10*(1), 14149. https://doi.org/10.1038/s41598-020-70941-y

Morgunova, E., & Taipale, J. (2017). Structural perspective of cooperative transcription factor binding. *Current Opinion in Structural Biology*, *47*, 1–8. https://doi.org/10.1016/j.sbi.2017.03.006

Mueller, D., Bach, C., Zeisig, D., Garcia-Cuellar, M.-P., Monroe, S., Sreekumar, A., Zhou, R., Nesvizhskii, A., Chinnaiyan, A., Hess, J. L., & Slany, R. K. (2007). A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. *Blood*, *110*(13), 4445–4454. https://doi.org/10.1182/blood-2007-05-090514

Müller, A. M., Medvinsky, A., Strouboulis, J., Grosveld, F., & Dzierzak, E. (1994). Development of hematopoietic stem cell activity in the mouse embryo. *Immunity*, *1*(4), 291–301. https://doi.org/10.1016/1074-7613(94)90081-7

Mundade, R., Ozer, H. G., Wei, H., Prabhu, L., & Lu, T. (2014). Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle*, *13*(18), 2847–2852. https://doi.org/10.4161/15384101.2014.949201

Muntean, A. G., & Hess, J. L. (2012). The pathogenesis of mixed-lineage leukemia. *Annual Review of Pathology*, *7*, 283–301. https://doi.org/10.1146/annurev-pathol-011811-132434

Murray, L. J., Bruno, E., Uchida, N., Hoffman, R., Nayar, R., Yeo, E. L., Schuh, A. C., & Sutherland, D. R. (1999). CD109 is expressed on a subpopulation of CD34+ cells enriched in hematopoietic stem and progenitor cells. *Experimental Hematology*, *27*(8), 1282–1294. https://doi.org/10.1016/s0301-472x(99)00071-5

Nabet, B., Roberts, J. M., Buckley, D. L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A. L., Erb, M. A., Lawlor, M. A., Souza, A., Scott, T. G., Vittori, S., Perry, J. A., Qi, J., Winter, G. E., Wong, K.-K., Gray, N. S., & Bradner, J. E. (2018). The dTAG system for immediate and target-specific protein degradation. *Nature Chemical Biology*, *14*(5), 431–441. https://doi.org/10.1038/s41589-018-0021-8

Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C. M., & Canaani, E. (2002). ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Molecular Cell*, *10*(5), 1119–1128. https://doi.org/10.1016/s1097-2765(02)00740-2

National Cancer Institute. (2022). SEER*Explorer: An interactive website for SEER cancer statistics [Internet] Surveillance Research Program, National Cancer Institute. Accessed on 26/09/2022.

Ng, C. E. L., Yokomizo, T., Yamashita, N., Cirovic, B., Jin, H., Wen, Z., Ito, Y., & Osato, M. (2010). A Runx1 intronic enhancer marks hemogenic endothelial cells and hematopoietic stem cells. *Stem Cells (Dayton, Ohio)*, *28*(10), 1869–1881. https://doi.org/10.1002/stem.507

Nichogiannopoulou, A., Trevisan, M., Neben, S., Friedrich, C., & Georgopoulos, K. (1999). Defects in hemopoietic stem cell activity in Ikaros mutant mice. *The Journal of Experimental Medicine*, *190*(9), 1201–1214. https://doi.org/10.1084/jem.190.9.1201

Nieminen, J. (1974). On the centrality in a graph. *Scandinavian Journal of Psychology*, *15*(4), 332–336. https://doi.org/10.1111/j.1467-9450.1974.tb00598.x

Niu, X., Wang, G., Wang, Y., Caldwell, J. T., Edwards, H., Xie, C., Taub, J. W., Li, C., Lin, H., & Ge, Y. (2014). Acute Myeloid Leukemia Cells Harboring MLL Fusion Genes or with the Acute Promyelocytic Leukemia Phenotype Are Sensitive to the Bcl-2-selective Inhibitor ABT-199. *Leukemia*, *28*(7), 1557–1560. https://doi.org/10.1038/leu.2014.72

Nordin, K., & LaBonne, C. (2014). Sox5 is a DNA binding co-factor for BMP R-Smads that directs target specificity during patterning of the early ectoderm. *Developmental cell*, *31*(3), 374–382. https://doi.org/10.1016/j.devcel.2014.10.003

North, T., Gu, T. L., Stacy, T., Wang, Q., Howard, L., Binder, M., Marín-Padilla, M., & Speck, N. A. (1999). Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development (Cambridge, England)*, *126*(11), 2563–2575. https://doi.org/10.1242/dev.126.11.2563

Nottingham, W. T., Jarratt, A., Burgess, M., Speck, C. L., Cheng, J.-F., Prabhakar, S., Rubin, E. M., Li, P.-S., Sloane-Stanley, J., Kong-a-San, J., & de Bruijn, M. F. T. R. (2007). Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. *Blood*, *110*(13), 4188–4197. https://doi.org/10.1182/blood-2007-07-100883

Nutt, S. L., Eberhard, D., Horcher, M., Rolink, A. G., & Busslinger, M. (2001). Pax5 determines the identity of B cells from the beginning to the end of B-lymphopoiesis. *International Reviews of Immunology*, *20*(1), 65–82. https://doi.org/10.3109/08830180109056723

O'Byrne, S., Elliott, N., Rice, S., Buck, G., Fordham, N., Garnett, C., Godfrey, L., Crump, N. T., Wright, G., Inglott, S., Hua, P., Psaila, B., Povinelli, B., Knapp, D. J. H. F., Agraz-Doblas, A., Bueno, C., Varela, I., Bennett, P., Koohy, H., … Roy, A. (2019). Discovery of a CD10-negative B-progenitor in human fetal life identifies unique ontogeny-related developmental programs. *Blood*, *134*(13), 1059–1071. https://doi.org/10.1182/blood.2019001289

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., & Nishimura, Y. (1994). Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell*, *79*(4), 639–648. https://doi.org/10.1016/0092-8674(94)90549-5

Ogawa, E., Inuzuka, M., Maruyama, M., Satake, M., Naito-Fujimoto, M., Ito, Y., & Shigesada, K. (1993). Molecular cloning and characterization of PEBP2 beta, the heterodimeric partner of a novel Drosophila runt-related DNA binding protein PEBP2 alpha. *Virology*, *194*(1), 314–331. https://doi.org/10.1006/viro.1993.1262

Ohmori, S., Moriguchi, T., Noguchi, Y., Ikeda, M., Kobayashi, K., Tomaru, N., Ishijima, Y., Ohneda, O., Yamamoto, M., & Ohneda, K. (2015). GATA2 is critical for the maintenance of cellular identity in differentiated mast cells derived from mouse bone marrow. *Blood*, *125*(21), 3306–3315. https://doi.org/10.1182/blood-2014-11-612465

Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V. M., Su, L., Xu, G., & Zhang, Y. (2005). hDOT1L links histone methylation to leukemogenesis. *Cell*, *121*(2), 167–178. https://doi.org/10.1016/j.cell.2005.02.020

Okuda, H., Kanai, A., Ito, S., Matsui, H., & Yokoyama, A. (2015). AF4 uses the SL1 components of RNAP1 machinery to initiate MLL fusion- and AEP-

dependent transcription. *Nature Communications*, *6*, 8869. https://doi.org/10.1038/ncomms9869

Okuda, T., van Deursen, J., Hiebert, S. W., Grosveld, G., & Downing, J. R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell*, *84*(2), 321–330. https://doi.org/10.1016/s0092-8674(00)80986-1

Oliveri, P., & Davidson, E. H. (2004). Gene regulatory network controlling embryonic specification in the sea urchin. *Current Opinion in Genetics & Development*, *14*(4), 351–360. https://doi.org/10.1016/j.gde.2004.06.004

Orlando, D. A., Chen, M. W., Brown, V. E., Solanki, S., Choi, Y. J., Olson, E. R., Fritz, C. C., Bradner, J. E., & Guenther, M. G. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Reports*, *9*(3), 1163–1170. https://doi.org/10.1016/j.celrep.2014.10.018

Ortt, K., Raveh, E., Gat, U., & Sinha, S. (2008). A chromatin immunoprecipitation screen in mouse keratinocytes reveals Runx1 as a direct transcriptional target of DeltaNp63. *Journal of Cellular Biochemistry*, *104*(4), 1204–1219. https://doi.org/10.1002/jcb.21700

Ottersbach, K. (2019). Endothelial-to-haematopoietic transition: An update on the process of making blood. *Biochemical Society Transactions*, *47*(2), 591–601. https://doi.org/10.1042/BST20180320

Owens, D. D. G., Anselmi, G., Oudelaar, A. M., Downes, D. J., Cavallo, A., Harman, J. R., Schwessinger, R., Bucakci, A., Greder, L., de Ornellas, S., Jeziorska, D., Telenius, J., Hughes, J. R., & de Bruijn, M. F. T. R. (2022). Dynamic Runx1 chromatin boundaries affect gene expression in hematopoietic development. *Nature Communications*, *13*(1), 773. https://doi.org/10.1038/s41467-022-28376-8

Padrón-Barthe, L., Temiño, S., Villa del Campo, C., Carramolino, L., Isern, J., & Torres, M. (2014). Clonal analysis identifies hemogenic endothelium as the source of the blood-endothelial common lineage in the mouse embryo. *Blood*, *124*(16), 2523–2532. https://doi.org/10.1182/blood-2013-12-545939

Palis, J., Robertson, S., Kennedy, M., Wall, C., & Keller, G. (1999). Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development (Cambridge, England)*, *126*(22), 5073–5084. https://doi.org/10.1242/dev.126.22.5073

Palis, J., & Yoder, M. C. (2001). Yolk-sac hematopoiesis: The first blood cells of mouse and man. *Experimental Hematology*, *29*(8), 927–936. https://doi.org/10.1016/S0301-472X(01)00669-5

Parslow, T. G., & Granner, D. K. (1983). Structure of a nuclease-sensitive region inside the immunoglobin kappa gene: Evidence for a role in gene regulation. *Nucleic Acids Research*, *11*(14), 4775–4792. https://doi.org/10.1093/nar/11.14.4775

Pearson, S., Cuvertino, S., Fleury, M., Lacaud, G., & Kouskoff, V. (2015). In Vivo Repopulating Activity Emerges at the Onset of Hematopoietic Specification during Embryonic Stem Cell Differentiation. *Stem Cell Reports*, *4*(3), 431–444. https://doi.org/10.1016/j.stemcr.2015.01.003

Perry, C., Eldor, A., & Soreq, H. (2002). Runx1/AML1 in leukemia: Disrupted association with diverse protein partners. *Leukemia research*, *26*(3). https://doi.org/10.1016/s0145-2126(01)00128-x

Peter, I. S., Faure, E., & Davidson, E. H. (2012). Predictive computation of genomic logic processing functions in embryonic development. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(41), 16434–16442. https://doi.org/10.1073/pnas.1207852109

Pieters, R., De Lorenzo, P., Ancliffe, P., Aversa, L. A., Brethon, B., Biondi, A., Campbell, M., Escherich, G., Ferster, A., Gardner, R. A., Kotecha, R. S., Lausen, B., Li, C. K., Locatelli, F., Attarbaschi, A., Peters, C., Rubnitz, J. E., Silverman, L. B., Stary, J., … Valsecchi, M. G. (2019). Outcome of Infants Younger Than 1 Year With Acute Lymphoblastic Leukemia Treated With the Interfant-06 Protocol: Results From an International Phase III Randomized Study. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *37*(25), 2246–2256. https://doi.org/10.1200/JCO.19.00261

Pinto do O, P., Kolterud, A., & Carlsson, L. (1998). Expression of the LIM-homeobox gene LH2 generates immortalized steel factor-dependent multipotent hematopoietic precursors. *The EMBO Journal*, *17*(19), 5744–5756. https://doi.org/10.1093/emboj/17.19.5744

Popescu, D.-M., Botting, R. A., Stephenson, E., Green, K., Webb, S., Jardine, L., Calderbank, E. F., Polanski, K., Goh, I., Efremova, M., Acres, M., Maunder, D., Vegh, P., Gitton, Y., Park, J.-E., Vento-Tormo, R., Miao, Z., Dixon, D., Rowell, R., … Haniffa, M. (2019). Decoding human fetal liver haematopoiesis. *Nature*, *574*(7778), 365–371. https://doi.org/10.1038/s41586-019-1652-y

Popovic, R., Riesbeck, L. E., Velu, C. S., Chaubey, A., Zhang, J., Achille, N. J., Erfurth, F. E., Eaton, K., Lu, J., Grimes, H. L., Chen, J., Rowley, J. D., & Zeleznik-Le, N. J. (2009). Regulation of mir-196b by MLL and its overexpression by MLL fusions contributes to immortalization. *Blood*, *113*(14), 3314–3322. https://doi.org/10.1182/blood-2008-04-154310

Prange, K. H. M., Mandoli, A., Kuznetsova, T., Wang, S.-Y., Sotoca, A. M., Marneth, A. E., van der Reijden, B. A., Stunnenberg, H. G., & Martens, J. H. A. (2017). MLL-AF9 and MLL-AF4 oncofusion proteins bind a distinct enhancer repertoire and target the RUNX1 program in 11q23 acute myeloid leukemia. *Oncogene*, *36*(23), 3346–3356. https://doi.org/10.1038/onc.2016.488

Pronk, C. J. H., Rossi, D. J., Månsson, R., Attema, J. L., Norddahl, G. L., Chan, C. K. F., Sigvardsson, M., Weissman, I. L., & Bryder, D. (2007). Elucidation of the Phenotypic, Functional, and Molecular Topography of a Myeloerythroid Progenitor Cell Hierarchy. *Cell Stem Cell*, *1*(4), 428–442. https://doi.org/10.1016/j.stem.2007.07.005

Puram, R. V., Kowalczyk, M. S., de Boer, C. G., Schneider, R. K., Miller, P. G., McConkey, M., Tothova, Z., Tejero, H., Heckl, D., Järås, M., Chen, M. C., Li, H., Tamayo, A., Cowley, G. S., Rozenblatt-Rosen, O., Al-Shahrour, F., Regev, A., & Ebert, B. L. (2016). Core circadian clock genes regulate leukemia stem cells in AML. *Cell*, *165*(2), 303–316. https://doi.org/10.1016/j.cell.2016.03.015

Quina, A. S., Buschbeck, M., & Di Croce, L. (2006). Chromatin structure and epigenetics. *Biochemical Pharmacology*, *72*(11), 1563–1569. https://doi.org/10.1016/j.bcp.2006.06.016

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2021). R: A Language and Environment for Statistical Computing.

Raffel, S., Falcone, M., Kneisel, N., Hansson, J., Wang, W., Lutz, C., Bullinger, L., Poschet, G., Nonnenmacher, Y., Barnert, A., Bahr, C., Zeisberger, P., Przybylla, A., Sohn, M., Tönjes, M., Erez, A., Adler, L., Jensen, P., Scholl, C., … Trumpp, A. (2017). BCAT1 restricts αKG levels in AML stem cells leading to IDHmut-like DNA hypermethylation. *Nature*, *551*(7680), 384–388. https://doi.org/10.1038/nature24294

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(Web Server issue), W160–W165. https://doi.org/10.1093/nar/gkw257

Reaman, G. H., Sposto, R., Sensel, M. G., Lange, B. J., Feusner, J. H., Heerema, N. A., Leonard, M., Holmes, E. J., Sather, H. N., Pendergrass, T. W., Johnstone, H. S., O'Brien, R. T., Steinherz, P. G., Zeltzer, P. M., Gaynon, P. S., Trigg, M. E., & Uckun, F. M. (1999). Treatment outcome and prognostic factors for infants with acute lymphoblastic leukemia treated on two consecutive trials of the Children's

Cancer Group. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *17*(2), 445–455. https://doi.org/10.1200/JCO.1999.17.2.445

Reiter, F., Wienerroither, S., & Stark, A. (2017). Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics & Development*, *43*, 73–81. https://doi.org/10.1016/j.gde.2016.12.007

Replogle, J. M., Saunders, R. A., Pogson, A. N., Hussmann, J. A., Lenail, A., Guna, A., Mascibroda, L., Wagner, E. J., Adelman, K., Lithwick-Yanai, G., Iremadze, N., Oberstrass, F., Lipson, D., Bonnar, J. L., Jost, M., Norman, T. M., & Weissman, J. S. (2022). Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell*, *185*(14), 2559–2575.e28. https://doi.org/10.1016/j.cell.2022.05.013

Reya, T., O'Riordan, M., Okamura, R., Devaney, E., Willert, K., Nusse, R., & Grosschedl, R. (2000). Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. *Immunity*, *13*(1), 15–24. https://doi.org/10.1016/s1074-7613(00)00004-2

Rice, S., Jackson, T., Crump, N. T., Fordham, N., Elliott, N., O'Byrne, S., Fanego, M. d. M. L., Addy, D., Crabb, T., Dryden, C., Inglott, S., Ladon, D., Wright, G., Bartram, J., Ancliff, P., Mead, A. J., Halsey, C., Roberts, I., Milne, T. A., & Roy, A. (2021). A human fetal liver-derived infant MLL-AF4 acute lymphoblastic leukemia model reveals a distinct fetal gene expression program. *Nature Communications*, *12*(1), 6905. https://doi.org/10.1038/s41467-021-27270-z

Rice, S., & Roy, A. (2020). MLL-rearranged infant leukaemia: A 'thorn in the side' of a remarkable success story. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1863*(8), 194564. https://doi.org/10.1016/j.bbagrm.2020.194564

Riddell, J., Gazit, R., Garrison, B. S., Guo, G., Saadatpour, A., Mandal, P. K., Ebina, W., Volchkov, P., Yuan, G.-C., Orkin, S. H., & Rossi, D. J. (2014). Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. *Cell*, *157*(3), 549–564. https://doi.org/10.1016/j.cell.2014.04.006

Ritz, C., Baty, F., Streibig, J. C., & Gerhard, D. (2015). Dose-Response Analysis Using R. *PLOS ONE*, *10*(12), e0146021. https://doi.org/10.1371/journal.pone.0146021

Robert-Moreno, A., Espinosa, L., de la Pompa, J. L., & Bigas, A. (2005). RBPjkappa-dependent Notch function regulates Gata2 and is essential for the formation of

intra-embryonic hematopoietic cells. *Development (Cambridge, England)*, *132*(5), 1117–1126. https://doi.org/10.1242/dev.01660

Robinson, B. W., Behling, K. C., Gupta, M., Zhang, A. Y., Moore, J. S., Bantly, A. D., Willman, C. L., Carroll, A. J., Adamson, P. C., Barrett, J. S., & Felix, C. A. (2008). Abundant anti-apoptotic BCL-2 is a molecular target in leukaemias with t(4;11) translocation. *British Journal of Haematology*, *141*(6), 827–839. https://doi.org/10.1111/j.1365-2141.2008.07100.x

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rodríguez-Martínez, J. A., Reinke, A. W., Bhimsaria, D., Keating, A. E., & Ansari, A. Z. (2017). Combinatorial bZIP dimers display complex DNA-binding specificity landscapes (M. R. Green, Ed.). *eLife*, *6*, e19272. https://doi.org/10.7554/eLife.19272

Roe, J.-S., & Vakoc, C. R. (2014). C/EBPα: Critical at the origin of leukemic transformation. *The Journal of Experimental Medicine*, *211*(1), 1–4. https://doi.org/10.1084/jem.20132530

Roeder, R. G., & Rutter, W. J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature*, *224*(5216), 234–237. https://doi.org/10.1038/224234a0

Rosenfeld, N., & Alon, U. (2003). Response Delays and the Structure of Transcription Networks. *Journal of Molecular Biology*, *329*(4), 645–654. https://doi.org/10.1016/S0022-2836(03)00506-0

Rothenberg, E. V., & Göttgens, B. (2021). How haematopoiesis research became a fertile ground for regulatory network biology as pioneered by Eric Davidson. *Current Opinion in Hematology*, *28*(1), 1–10. https://doi.org/10.1097/MOH.0000000000000628

Roy, A., Wang, G., Iskander, D., O'Byrne, S., Elliott, N., O'Sullivan, J., Buck, G., Heuston, E. F., Wen, W. X., Meira, A. R., Hua, P., Karadimitris, A., Mead, A. J., Bodine, D. M., Roberts, I., Psaila, B., & Thongjuea, S. (2021). Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development. *Cell Reports*, *36*(11), 109698. https://doi.org/10.1016/j.celrep.2021.109698

Roychoudhuri, R., Clever, D., Li, P., Wakabayashi, Y., Quinn, K. M., Klebanoff, C. A., Ji, Y., Sukumar, M., Eil, R. L., Yu, Z., Spolski, R., Palmer, D. C.,

Pan, J. H., Patel, S. J., Macallan, D. C., Fabozzi, G., Shih, H.-Y., Kanno, Y., Muto, A., ... Restifo, N. P. (2016). BACH2 regulates CD8(+) T cell differentiation by controlling access of AP-1 factors to enhancers. *Nature Immunology*, *17*(7), 851–860. https://doi.org/10.1038/ni.3441

Rybtsov, S., Batsivari, A., Bilotkach, K., Paruzina, D., Senserrich, J., Nerushev, O., & Medvinsky, A. (2014). Tracing the origin of the HSC hierarchy reveals an SCF-dependent, IL-3-independent CD43(-) embryonic precursor. *Stem Cell Reports*, *3*(3), 489–501. https://doi.org/10.1016/j.stemcr.2014.07.009

Rybtsov, S., Sobiesiak, M., Taoudi, S., Souilhol, C., Senserrich, J., Liakhovitskaia, A., Ivanovs, A., Frampton, J., Zhao, S., & Medvinsky, A. (2011). Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. *The Journal of Experimental Medicine*, *208*(6), 1305–1315. https://doi.org/10.1084/jem.20102419

Samokhvalov, I. M., Samokhvalova, N. I., & Nishikawa, S.-i. (2007). Cell tracing shows the contribution of the yolk sac to adult haematopoiesis. *Nature*, *446*(7139), 1056–1061. https://doi.org/10.1038/nature05725

Sandberg, M. L., Sutton, S. E., Pletcher, M. T., Wiltshire, T., Tarantino, L. M., Hogenesch, J. B., & Cooke, M. P. (2005). C-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation. *Developmental Cell*, *8*(2), 153–166. https://doi.org/10.1016/j.devcel.2004.12.015

Sanyal, A., Lajoie, B. R., Jain, G., & Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature*, *489*(7414), 109–113. https://doi.org/10.1038/nature11279

Satpathy, A. T., Granja, J. M., Yost, K. E., Qi, Y., Meschi, F., McDermott, G. P., Olsen, B. N., Mumbach, M. R., Pierce, S. E., Corces, M. R., Shah, P., Bell, J. C., Jhutty, D., Nemec, C. M., Wang, J., Wang, L., Yin, Y., Giresi, P. G., Chang, A. L. S., ... Chang, H. Y. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nature biotechnology*, *37*(8), 925–936. https://doi.org/10.1038/s41587-019-0206-z

Schütte, J., Wang, H., Antoniou, S., Jarratt, A., Wilson, N. K., Riepsaame, J., Calero-Nieto, F. J., Moignard, V., Basilico, S., Kinston, S. J., Hannah, R. L., Chan, M. C., Nürnberg, S. T., Ouwehand, W. H., Bonzanni, N., de Bruijn, M. F., & Göttgens, B. (2016). An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability (A. J. Wagers, Ed.). *eLife*, *5*, e11469. https://doi.org/10.7554/eLife.11469

Sciammas, R., Li, Y., Warmflash, A., Song, Y., Dinner, A. R., & Singh, H. (2011). An incoherent regulatory network architecture that orchestrates B cell diversification

in response to antigen signaling. *Molecular Systems Biology*, *7*, 495. https://doi.org/10.1038/msb.2011.25

Scripture-Adams, D. D., Damle, S. S., Li, L., Elihu, K. J., Qin, S., Arias, A. M., Butler, R. R., Champhekar, A., Zhang, J. A., & Rothenberg, E. V. (2014). GATA-3 dose-dependent checkpoints in early T cell commitment. *Journal of immunology (Baltimore, Md. : 1950)*, *193*(7), 3470–3491. https://doi.org/10.4049/jimmunol.1301663

Segert, J. A., Gisselbrecht, S. S., & Bulyk, M. L. (2021). Transcriptional Silencers: Driving Gene Expression with the Brakes On. *Trends in Genetics*, *37*(6), 514–527. https://doi.org/10.1016/j.tig.2021.02.002

Serin, E. A. R., Nijveen, H., Hilhorst, H. W. M., & Ligterink, W. (2016). Learning from Co-expression Networks: Possibilities and Challenges. *Frontiers in Plant Science*, *7*.

Severin, J., Waterhouse, A. M., Kawaji, H., Lassmann, T., van Nimwegen, E., Balwierz, P. J., de Hoon, M. J., Hume, D. A., Carninci, P., Hayashizaki, Y., Suzuki, H., Daub, C. O., & Forrest, A. R. (2009). FANTOM4 EdgeExpressDB: An integrated database of promoters, genes, microRNAs, expression dynamics and regulatory interactions. *Genome Biology*, *10*(4), R39. https://doi.org/10.1186/gb-2009-10-4-r39

Sheffield, N. C., Thurman, R. E., Song, L., Safi, A., Stamatoyannopoulos, J. A., Lenhard, B., Crawford, G. E., & Furey, T. S. (2013). Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Research*, *23*(5), 777–788. https://doi.org/10.1101/gr.152140.112

Shen-Orr, S. S., Milo, R., Mangan, S., & Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genetics*, *31*(1), 64–68. https://doi.org/10.1038/ng881

Shi, J., Whyte, W. A., Zepeda-Mendoza, C. J., Milazzo, J. P., Shen, C., Roe, J.-S., Minder, J. L., Mercan, F., Wang, E., Eckersley-Maslin, M. A., Campbell, A. E., Kawaoka, S., Shareef, S., Zhu, Z., Kendall, J., Muhar, M., Haslinger, C., Yu, M., Roeder, R. G., … Vakoc, C. R. (2013). Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes & Development*, *27*(24), 2648–2662. https://doi.org/10.1101/gad.232710.113

Shin, B., Hosokawa, H., Romero-Wolf, M., Zhou, W., Masuhara, K., Tobin, V. R., Levanon, D., Groner, Y., & Rothenberg, E. V. (2021). Runx1 and Runx3 drive progenitor to T-lineage transcriptome conversion in mouse T cell commitment via

dynamic genomic site switching. *Proceedings of the National Academy of Sciences*, *118*(4), e2019655118. https://doi.org/10.1073/pnas.2019655118

Shin, H. Y., Willi, M., HyunYoo, K., Zeng, X., Wang, C., Metser, G., & Hennighausen, L. (2016). Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nature Genetics*, *48*(8), 904–911. https://doi.org/10.1038/ng.3606

Shmulevich, I., Dougherty, E. R., Kim, S., & Zhang, W. (2002). Probabilistic Boolean Networks: A rule-based uncertainty model for gene regulatory networks. *Bioinformatics (Oxford, England)*, *18*(2), 261–274. https://doi.org/10.1093/bioinformatics/18.2.261

Siggers, T., Duyzend, M. H., Reddy, J., Khan, S., & Bulyk, M. L. (2011). Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. *Molecular Systems Biology*, *7*, 555. https://doi.org/10.1038/msb.2011.89

Simeoni, F., Romero-Camarero, I., Camera, F., Amaral, F. M. R., Sinclair, O. J., Papachristou, E. K., Spencer, G. J., Lie-A-Ling, M., Lacaud, G., Wiseman, D. H., Carroll, J. S., & Somervaille, T. C. P. (2021). Enhancer recruitment of transcription repressors RUNX1 and TLE3 by mis-expressed FOXC1 blocks differentiation in acute myeloid leukemia. *Cell Reports*, *36*(12), 109725. https://doi.org/10.1016/j.celrep.2021.109725

Simões-Costa, M., & Bronner, M. E. (2015). Establishing neural crest identity: A gene regulatory recipe. *Development (Cambridge, England)*, *142*(2), 242–257. https://doi.org/10.1242/dev.105445

Singh, R., Letai, A., & Sarosiek, K. (2019). Regulation of apoptosis in health and disease: The balancing act of BCL-2 family proteins. *Nature Reviews Molecular Cell Biology*, *20*(3), 175–193. https://doi.org/10.1038/s41580-018-0089-8

Slany, R. K. (2020). MLL fusion proteins and transcriptional control. *Biochimica Et Biophysica Acta. Gene Regulatory Mechanisms*, *1863*(3), 194503. https://doi.org/10.1016/j.bbagrm.2020.194503

Smith, E., Lin, C., & Shilatifard, A. (2011). The super elongation complex (SEC) and MLL in development and disease. *Genes & Development*, *25*(7), 661–672. https://doi.org/10.1101/gad.2015411

Solaimani Kartalaei, P., Yamada-Inagawa, T., Vink, C. S., de Pater, E., van der Linden, R., Marks-Bluth, J., van der Sloot, A., van den Hout, M., Yokomizo, T., van Schaick-Solernó, M. L., Delwel, R., Pimanda, J. E., van IJcken, W. F. J., & Dzierzak, E. (2015). Whole-transcriptome analysis of endothelial to hematopoietic stem cell transition reveals a requirement for Gpr56 in HSC generation. *The*

Journal of Experimental Medicine, *212*(1), 93–106. https://doi.org/10.1084/jem.20140767

Søndergaard, E., Rauch, A., Michaut, M., Rapin, N., Rehn, M., Wilhelmson, A. S., Camponeschi, A., Hasemann, M. S., Bagger, F. O., Jendholm, J., Knudsen, K. J., Mandrup, S., Mårtensson, I.-L., & Porse, B. T. (2019). ERG Controls B Cell Development by Promoting Igh V-to-DJ Recombination. *Cell Reports*, *29*(9), 2756–2769.e6. https://doi.org/10.1016/j.celrep.2019.10.098

Soto-Feliciano, Y. M., Sánchez-Rivera, F. J., Perner, F., Barrows, D. W., Kastenhuber, E. R., Ho, Y.-J., Carroll, T., Xiong, Y., Anand, D., Soshnev, A., Gates, L., Beytagh, M. C., Cheon, D., Gu, S., Liu, X. S., Krivtsov, A. V., Meneses, M., Stanchina, E. d., Stone, R. M., … Allis, C. D. (2022). A molecular switch between mammalian MLL complexes dictates response to Menin-MLL inhibition [Section: New Results], 2021.10.22.465184. https://doi.org/10.1101/2021.10.22.465184

Souilhol, C., Lendinez, J. G., Rybtsov, S., Murphy, F., Wilson, H., Hills, D., Batsivari, A., Binagui-Casas, A., McGarvey, A. C., MacDonald, H. R., Kageyama, R., Siebel, C., Zhao, S., & Medvinsky, A. (2016). Developing HSCs become Notch independent by the end of maturation in the AGM region. *Blood*, *128*(12), 1567–1577. https://doi.org/10.1182/blood-2016-03-708164

Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics*, *13*(9), 613–626. https://doi.org/10.1038/nrg3207

Stanchi, F., Corso, V., Scannapieco, P., Ievolella, C., Negrisolo, E., Tiso, N., Lanfranchi, G., & Valle, G. (2000). TUBA8: A new tissue-specific isoform of alpha-tubulin that is highly conserved in human and mouse. *Biochemical and Biophysical Research Communications*, *270*(3), 1111–1118. https://doi.org/10.1006/bbrc.2000.2571

Steger, D. J., Lefterova, M. I., Ying, L., Stonestrom, A. J., Schupp, M., Zhuo, D., Vakoc, A. L., Kim, J.-E., Chen, J., Lazar, M. A., Blobel, G. A., & Vakoc, C. R. (2008). DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Molecular and Cellular Biology*, *28*(8), 2825–2839. https://doi.org/10.1128/MCB.02076-07

Stormo, G. D. (2000). DNA binding sites: Representation and discovery. *Bioinformatics*, *16*(1), 16–23. https://doi.org/10.1093/bioinformatics/16.1.16

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E., & Dudoit, S. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, *19*(1), 477. https://doi.org/10.1186/s12864-018-4772-0

Sturm, R. A., Das, G., & Herr, W. (1988). The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. *Genes & Development*, *2*(12A), 1582–1599. https://doi.org/10.1101/gad.2.12a.1582

Suo, C., Dann, E., Goh, I., Jardine, L., Kleshchevnikov, V., Park, J.-E., Botting, R. A., Stephenson, E., Engelbert, J., Tuong, Z. K., Polanski, K., Yayon, N., Xu, C., Suchanek, O., Elmentaite, R., Domínguez Conde, C., He, P., Pritchard, S., Miah, M., … Teichmann, S. A. (2022). Mapping the developing human immune system across organs. *Science*, *376*(6597), eabo0510. https://doi.org/10.1126/science.abo0510

Swiers, G., Baumann, C., O'Rourke, J., Giannoulatou, E., Taylor, S., Joshi, A., Moignard, V., Pina, C., Bee, T., Kokkaliaris, K. D., Yoshimoto, M., Yoder, M. C., Frampton, J., Schroeder, T., Enver, T., Göttgens, B., & de Bruijn, M. F. T. R. (2013a). Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nature Communications*, *4*, 2924. https://doi.org/10.1038/ncomms3924

Swiers, G., Rode, C., Azzoni, E., & de Bruijn, M. F. T. R. (2013b). A short history of hemogenic endothelium. *Blood Cells, Molecules & Diseases*, *51*(4), 206–212. https://doi.org/10.1016/j.bcmd.2013.09.005

Takahashi, H., Parmely, T. J., Sato, S., Tomomori-Sato, C., Banks, C. A. S., Kong, S. E., Szutorisz, H., Swanson, S. K., Martin-Brown, S., Washburn, M. P., Florens, L., Seidel, C., Lin, C., Smith, E. R., Shilatifard, A., Conaway, R. C., & Conaway, J. W. (2011). Human Mediator Subunit Med26 Functions As A Docking Site For Transcription Elongation Factors. *Cell*, *146*(1), 92–104. https://doi.org/10.1016/j.cell.2011.06.005

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Takahashi, S., & Yokoyama, A. (2020). The molecular functions of common and atypical MLL fusion protein complexes. *Biochimica Et Biophysica Acta. Gene Regulatory Mechanisms*, *1863*(7), 194548. https://doi.org/10.1016/j.bbagrm.2020.194548

Taoudi, S., Gonneau, C., Moore, K., Sheridan, J. M., Blackburn, C. C., Taylor, E., & Medvinsky, A. (2008). Extensive Hematopoietic Stem Cell Generation in the AGM Region via Maturation of VE-Cadherin+CD45+ Pre-Definitive HSCs. *Cell Stem Cell*, *3*(1), 99–108. https://doi.org/10.1016/j.stem.2008.06.004

Tegnér, J., & Björkegren, J. (2007). Perturbations to uncover gene networks. *Trends in genetics: TIG*, *23*(1), 34–41. https://doi.org/10.1016/j.tig.2006.11.003

Telenius, J., Consortium, T. W., & Hughes, J. R. (2018). NGseqBasic - a single-command UNIX tool for ATAC-seq, DNaseI-seq, Cut-and-Run, and ChIP-seq data mapping, high-resolution visualisation, and quality control. *bioRxiv*, 393413. https://doi.org/10.1101/393413

The Cancer Genome Atlas Research Network. (2013). Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *https://doi.org/10.1056/NEJMoa1301689*, (368), 2059–2074. https://doi.org/10.1056/NEJMoa1301689

The FANTOM Consortium and the Riken Omics Science Center. (2009). The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature Genetics*, *41*(5), 553–562. https://doi.org/10.1038/ng.375

Theisen, M., Stief, A., & Sippel, A. E. (1986). The lysozyme enhancer: Cell-specific activation of the chicken lysozyme gene by a far-upstream DNA element. *The EMBO journal*, *5*(4), 719–724. https://doi.org/10.1002/j.1460-2075.1986.tb04273.x

Thomas, M., Gessner, A., Vornlocher, H.-P., Hadwiger, P., Greil, J., & Heidenreich, O. (2005). Targeting MLL-AF4 with short interfering RNAs inhibits clonogenicity and engraftment of t(4;11)-positive human leukemic cells. *Blood*, *106*(10), 3559–3566. https://doi.org/10.1182/blood-2005-03-1283

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., & Narechania, A. (2003). PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Research*, *13*(9), 2129–2141. https://doi.org/10.1101/gr.772403

Thomas, R. (1973). Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, *42*(3), 563–585. https://doi.org/10.1016/0022-5193(73)90247-6

Thoms, J. A. I., Beck, D., & Pimanda, J. E. (2019). Transcriptional networks in acute myeloid leukemia. *Genes, Chromosomes & Cancer*. https://doi.org/10.1002/gcc.22794

Tomizawa, D., Koh, K., Sato, T., Kinukawa, N., Morimoto, A., Isoyama, K., Kosaka, Y., Oda, T., Oda, M., Hayashi, Y., Eguchi, M., Horibe, K., Nakahata, T., Mizutani, S., & Ishii, E. (2007). Outcome of risk-based therapy for infant acute lymphoblastic leukemia with or without an MLL gene rearrangement, with emphasis on late effects: A final report of two consecutive studies, MLL96 and MLL98, of the Japan Infant Leukemia Study Group. *Leukemia*, *21*(11), 2258–2263. https://doi.org/10.1038/sj.leu.2404903

Tsai, F. Y., Keller, G., Kuo, F. C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F. W., & Orkin, S. H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*, *371*(6494), 221–226. https://doi.org/10.1038/371221a0

Tsuzuki, S., Taguchi, O., & Seto, M. (2011). Promotion and maintenance of leukemia by ERG. *Blood*, *117*(14), 3858–3868. https://doi.org/10.1182/blood-2010-11-320515

Tzelepis, K., Koike-Yusa, H., De Braekeleer, E., Li, Y., Metzakopian, E., Dovey, O. M., Mupo, A., Grinkevich, V., Li, M., Mazan, M., Gozdecka, M., Ohnishi, S., Cooper, J., Patel, M., McKerrell, T., Chen, B., Domingues, A. F., Gallipoli, P., Teichmann, S., … Yusa, K. (2016). A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Reports*, *17*(4), 1193–1205. https://doi.org/10.1016/j.celrep.2016.09.079

Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., Verbeiren, T., De Maeyer, D., Reumers, J., Saeys, Y., & Aerts, S. (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nature Protocols*, *15*(7), 2247–2276. https://doi.org/10.1038/s41596-020-0336-2

Van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., Dudoit, S., & Clement, L. (2020). Trajectory-based differential expression analysis for single-cell sequencing data. *Nature Communications*, *11*(1), 1201. https://doi.org/10.1038/s41467-020-14766-3

Vandenberg, C. J., & Cory, S. (2013). ABT-199, a new Bcl-2–specific BH3 mimetic, has in vivo efficacy against aggressive Myc-driven mouse lymphomas without provoking thrombocytopenia. *Blood*, *121*(12), 2285–2288. https://doi.org/10.1182/blood-2013-01-475855

Vandewalle, C., Comijn, J., De Craene, B., Vermassen, P., Bruyneel, E., Andersen, H., Tulchinsky, E., Van Roy, F., & Berx, G. (2005). SIP1/ZEB2 induces EMT by repressing genes of different epithelial cell-cell junctions. *Nucleic Acids Research*, *33*(20), 6566–6578. https://doi.org/10.1093/nar/gki965

van Oevelen, C., Collombet, S., Vicent, G., Hoogenkamp, M., Lepoivre, C., Badeaux, A., Bussmann, L., Sardina, J. L., Thieffry, D., Beato, M., Shi, Y., Bonifer, C., & Graf, T. (2015). C/EBPα Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. *Stem Cell Reports*, *5*(2), 232–247. https://doi.org/10.1016/j.stemcr.2015.06.007

Vargel, Ö., Zhang, Y., Kosim, K., Ganter, K., Foehr, S., Mardenborough, Y., Shvartsman, M., Enright, A. J., Krijgsveld, J., & Lancrin, C. (2016). Activation

of the TGFβ pathway impairs endothelial to haematopoietic transition. *Scientific Reports*, *6*, 21518. https://doi.org/10.1038/srep21518

Venolia, L., & Gartler, S. M. (1983). Comparison of transformation efficiency of human active and inactive X-chromosomal DNA. *Nature*, *302*(5903), 82–83. https://doi.org/10.1038/302082a0

Vignali, R., & Marracci, S. (2020). HMGA Genes and Proteins in Development and Evolution. *International Journal of Molecular Sciences*, *21*(2), 654. https://doi.org/10.3390/ijms21020654

Vink, C. S., Calero-Nieto, F. J., Wang, X., Maglitto, A., Mariani, S. A., Jawaid, W., Göttgens, B., & Dzierzak, E. (2020). Iterative Single-Cell Analyses Define the Transcriptome of the First Functional Hematopoietic Stem Cells. *Cell Reports*, *31*(6), 107627. https://doi.org/10.1016/j.celrep.2020.107627

Wade, J. T. (2015). Mapping Transcription Regulatory Networks with ChIP-seq and RNA-seq. *Advances in Experimental Medicine and Biology*, *883*, 119–134. https://doi.org/10.1007/978-3-319-23603-2_7

Wang, H., Zang, C., Taing, L., Arnett, K. L., Wong, Y. J., Pear, W. S., Blacklow, S. C., Liu, X. S., & Aster, J. C. (2014). NOTCH1-RBPJ complexes drive target gene expression through dynamic interactions with superenhancers. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(2), 705–710. https://doi.org/10.1073/pnas.1315023111

Wang, L., Brugge, J. S., & Janes, K. A. (2011). Intersection of FOXO- and RUNX1-mediated gene expression programs in single breast epithelial cells during morphogenesis and tumor progression. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(40), E803–E812. https://doi.org/10.1073/pnas.1103423108

Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A. H., & Speck, N. A. (1996a). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(8), 3444–3449. https://doi.org/10.1073/pnas.93.8.3444

Wang, Q., Stacy, T., Miller, J. D., Lewis, A. F., Gu, T. L., Huang, X., Bushweller, J. H., Bories, J. C., Alt, F. W., Ryan, G., Liu, P. P., Wynshaw-Boris, A., Binder, M., Marín-Padilla, M., Sharpe, A. H., & Speck, N. A. (1996b). The CBFbeta subunit is essential for CBFalpha2 (AML1) function in vivo. *Cell*, *87*(4), 697–708. https://doi.org/10.1016/s0092-8674(00)81389-6

Wang, W., Schwemmers, S., Hexner, E. O., & Pahl, H. L. (2010). AML1 is overexpressed in patients with myeloproliferative neoplasms and mediates JAK2V617F-independent overexpression of NF-E2. *Blood*, *116*(2), 254–266. https://doi.org/10.1182/blood-2009-11-254664

Wang, X., Angelis, N., & Thein, S. L. (2018). MYB - A regulatory factor in hematopoiesis. *Gene*, *665*, 6–17. https://doi.org/10.1016/j.gene.2018.04.065

Watson, J. R., Owen, D., & Mott, H. R. (2016). Cdc42 in actin dynamics: An ordered pathway governed by complex equilibria and directional effector handover. *Small GTPases*, *8*(4), 237–244. https://doi.org/10.1080/21541248.2016.1215657

Weirauch, M. T., & Hughes, T. (2011). A Catalogue of Eukaryotic Transcription Factor Types, Their Evolutionary Origin, and Species Distribution. In T. R. Hughes (Ed.), *A Handbook of Transcription Factors* (pp. 25–73). Springer Netherlands. https://doi.org/10.1007/978-90-481-9069-0_3

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., & Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, *153*(2), 307–319. https://doi.org/10.1016/j.cell.2013.03.035

Wilkinson, A. C., Ballabio, E., Geng, H., North, P., Tapia, M., Kerry, J., Biswas, D., Roeder, R. G., Allis, C. D., Melnick, A., de Bruijn, M. F., & Milne, T. A. (2013). RUNX1 Is a Key Target in t(4;11) Leukemias that Contributes to Gene Activation through an AF4-MLL Complex Interaction. *Cell Reports*, *3*(1), 116–127. https://doi.org/10.1016/j.celrep.2012.12.016

Williams, R. M., Candido-Ferreira, I., Repapi, E., Gavriouchkina, D., Senanayake, U., Ling, I. T. C., Telenius, J., Taylor, S., Hughes, J., & Sauka-Spengler, T. (2019). Reconstruction of the Global Neural Crest Gene Regulatory Network In Vivo. *Developmental Cell*, *51*(2), 255–276.e7. https://doi.org/10.1016/j.devcel.2019.10.003

Wilson, N. K., Schoenfelder, S., Hannah, R., Sánchez Castillo, M., Schütte, J., Ladopoulos, V., Mitchelmore, J., Goode, D. K., Calero-Nieto, F. J., Moignard, V., Wilkinson, A. C., Jimenez-Madrid, I., Kinston, S., Spivakov, M., Fraser, P., & Göttgens, B. (2016). Integrated genome-scale analysis of the transcriptional regulatory landscape in a blood stem/progenitor cell model. *Blood*, *127*(13), e12–23. https://doi.org/10.1182/blood-2015-10-677393

Wilson, N. K., Timms, R. T., Kinston, S. J., Cheng, Y.-H., Oram, S. H., Landry, J.-R., Mullender, J., Ottersbach, K., & Gottgens, B. (2010). Gfi1 Expression Is Controlled by Five Distinct Regulatory Regions Spread over 100 Kilobases, with Scl/Tal1, Gata2, PU.1, Erg, Meis1, and Runx1 Acting as Upstream Regulators in

Early Hematopoietic Cells. *Molecular and Cellular Biology*, *30*(15), 3853–3863. https://doi.org/10.1128/MCB.00032-10

Winters, A. C., & Bernt, K. M. (2017). MLL-Rearranged Leukemias—An Update on Science and Clinical Approaches. *Frontiers in Pediatrics*, *5*, 4. https://doi.org/10.3389/fped.2017.00004

Woll, P. S., Grzywacz, B., Tian, X., Marcus, R. K., Knorr, D. A., Verneris, M. R., & Kaufman, D. S. (2009). Human embryonic stem cells differentiate into a homogeneous population of natural killer cells with potent in vivo antitumor activity. *Blood*, *113*(24), 6094–6101. https://doi.org/10.1182/blood-2008-06-165225

Xiang, G., Keller, C. A., Heuston, E., Giardine, B. M., An, L., Wixom, A. Q., Miller, A., Cockburn, A., Sauria, M. E. G., Weaver, K., Lichtenberg, J., Göttgens, B., Li, Q., Bodine, D., Mahony, S., Taylor, J., Blobel, G. A., Weiss, M. J., Cheng, Y., … Hardison, R. C. (2020). An integrative view of the regulatory and transcriptional landscapes in mouse hematopoiesis. *Genome Research*, *30*(3), 472–484. https://doi.org/10.1101/gr.255760.119

Xiang, W., Peng, T., Ming, Y., Li, S., Wang, K., Wang, H., Chen, L., & Zhou, J. (2019). Myristoylated alanine rich protein kinase C substrate is a potential cancer prognostic factor that regulates cell migration and invasion in glioblastoma. *Oncology Reports*, *41*(4), 2464–2470. https://doi.org/10.3892/or.2019.7009

Xie, H., Ye, M., Feng, R., & Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. *Cell*, *117*(5), 663–676. https://doi.org/10.1016/s0092-8674(04)00419-2

Yamagata, T., Maki, K., & Mitani, K. (2005). Runx1/AML1 in normal and abnormal hematopoiesis. *International journal of hematology*, *82*(1). https://doi.org/10.1532/IJH97.05075

Yáñez-Cuna, J. O., Dinh, H. Q., Kvon, E. Z., Shlyueva, D., & Stark, A. (2012). Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding. *Genome Research*, *22*(10), 2018–2030. https://doi.org/10.1101/gr.132811.111

Yang, Z., Yik, J. H. N., Chen, R., He, N., Jang, M. K., Ozato, K., & Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Molecular Cell*, *19*(4), 535–545. https://doi.org/10.1016/j.molcel.2005.06.029

Yokomizo, T., Yamada-Inagawa, T., Yzaguirre, A. D., Chen, M. J., Speck, N. A., & Dzierzak, E. (2012). Whole-mount three-dimensional imaging of internally

localized immunostained cells within mouse embryos. *Nature protocols*, *7*(3), 421–431. https://doi.org/10.1038/nprot.2011.441

Yokoyama, A., & Cleary, M. L. (2008). Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer cell*, *14*(1), 36–46. https://doi.org/10.1016/j.ccr.2008.05.003

Yokoyama, A., Lin, M., Naresh, A., Kitabayashi, I., & Cleary, M. L. (2010). A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. *Cancer Cell*, *17*(2), 198–212. https://doi.org/10.1016/j.ccr.2009.12.040

Yoshimoto, M., Montecino-Rodriguez, E., Ferkowicz, M. J., Porayette, P., Shelley, W. C., Conway, S. J., Dorshkind, K., & Yoder, M. C. (2011). Embryonic day 9 yolk sac and intra-embryonic hemogenic endothelium independently generate a B-1 and marginal zone progenitor lacking B-2 potential. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1468–1473. https://doi.org/10.1073/pnas.1015841108

Yoshimoto, M., Porayette, P., Glosson, N. L., Conway, S. J., Carlesso, N., Cardoso, A. A., Kaplan, M. H., & Yoder, M. C. (2012). Autonomous murine T-cell progenitor production in the extra-embryonic yolk sac before HSC emergence. *Blood*, *119*(24), 5706–5714. https://doi.org/10.1182/blood-2011-12-397489

Yuh, C. H., Bolouri, H., & Davidson, E. H. (1998). Genomic cis-regulatory logic: Experimental and computational analysis of a sea urchin gene. *Science (New York, N.Y.)*, *279*(5358), 1896–1902. https://doi.org/10.1126/science.279.5358.1896

Yuh, C. H., Ransick, A., Martinez, P., Britten, R. J., & Davidson, E. H. (1994). Complexity and organization of DNA-protein interactions in the 5'-regulatory region of an endoderm-specific marker gene in the sea urchin embryo. *Mechanisms of Development*, *47*(2), 165–186. https://doi.org/10.1016/0925-4773(94)90088-4

Zaidan, N., Nitsche, L., Diamanti, E., Hannah, R., Fidanza, A., Wilson, N. K., Forrester, L. M., Göttgens, B., & Ottersbach, K. (2022). Endothelial-specific Gata3 expression is required for hematopoietic stem cell generation. *Stem Cell Reports*, S2213–6711(22)00324–1. https://doi.org/10.1016/j.stemcr.2022.06.008

Zaidi, S. K., Sullivan, A. J., van Wijnen, A. J., Stein, J. L., Stein, G. S., & Lian, J. B. (2002). Integration of Runx and Smad regulatory signals at transcriptionally active subnuclear sites. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 8048–8053. https://doi.org/10.1073/pnas.112664499

Zaret, K. S. (2020). Pioneer Transcription Factors Initiating Gene Network Changes. *Annual review of genetics*, *54*, 367–385. https://doi.org/10.1146/annurev-genet-030220-015007

Zaret, K. S., & Mango, S. (2016). Pioneer Transcription Factors, Chromatin Dynamics, and Cell Fate Control. *Current opinion in genetics & development*, *37*, 76–81. https://doi.org/10.1016/j.gde.2015.12.003

Zeleznik-Le, N. J., Harden, A. M., & Rowley, J. D. (1994). 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(22), 10610–10614. https://doi.org/10.1073/pnas.91.22.10610

Zeng, Y., He, J., Bai, Z., Li, Z., Gong, Y., Liu, C., Ni, Y., Du, J., Ma, C., Bian, L., Lan, Y., & Liu, B. (2019). Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. *Cell Research*, *29*(11), 881–894. https://doi.org/10.1038/s41422-019-0228-6

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., … Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8*(1), 14049. https://doi.org/10.1038/ncomms14049

Zhou, B. O., Ding, L., & Morrison, S. J. (2015). Hematopoietic stem and progenitor cells regulate the regeneration of their niche by secreting Angiopoietin-1. *eLife*, *4*, e05521. https://doi.org/10.7554/eLife.05521

Zhou, N., Gutierrez-Uzquiza, A., Zheng, X. Y., Chang, R., Vogl, D. T., Garfall, A. L., Bernabei, L., Saraf, A., Florens, L., Washburn, M. P., Illendula, A., Bushweller, J. H., & Busino, L. (2019). RUNX proteins desensitize multiple myeloma to lenalidomide via protecting IKZFs from degradation. *Leukemia*, *33*(8), 2006–2021. https://doi.org/10.1038/s41375-019-0403-2

Zhou, W., Gao, F., Romero-Wolf, M., Jo, S., & Rothenberg, E. V. (2022). Single-cell deletion analyses show control of pro–T cell developmental speed and pathways by Tcf7, Spi1, Gata3, Bcl11a, Erg, and Bcl11b. *Science Immunology*, *7*(71), eabm1920. https://doi.org/10.1126/sciimmunol.abm1920

Zhu, L., Li, Q., Wong, S. H., Huang, M., Klein, B. J., Shen, J., Ikenouye, L., Onishi, M., Schneidawind, D., Buechele, C., Hansen, L., Duque-Afonso, J., Zhu, F., Martin, G. M., Gozani, O., Majeti, R., Kutateladze, T. G., & Cleary, M. L. (2016). ASH1L Links Histone H3 Lysine 36 di-methylation to MLL Leukemia. *Cancer discovery*, *6*(7), 770–783. https://doi.org/10.1158/2159-8290.CD-16-0058

Zhu, Q., Gao, P., Tober, J., Bennett, L., Chen, C., Uzun, Y., Li, Y., Howell, E. D., Mumau, M., Yu, W., He, B., Speck, N. A., & Tan, K. (2020). Developmental trajectory of prehematopoietic stem cell formation from endothelium. *Blood*, *136*(7), 845–856. https://doi.org/10.1182/blood.2020004801

Zi, J., Han, Q., Gu, S., McGrath, M., Kane, S., Song, C., & Ge, Z. (2020). Targeting NAT10 Induces Apoptosis Associated With Enhancing Endoplasmic Reticulum Stress in Acute Myeloid Leukemia Cells. *Frontiers in Oncology*, *10*, 598107. https://doi.org/10.3389/fonc.2020.598107

Zovein, A. C., Hofmann, J. J., Lynch, M., French, W. J., Turlo, K. A., Yang, Y., Becker, M. S., Zanetta, L., Dejana, E., Gasson, J. C., Tallquist, M. D., & Iruela-Arispe, M. L. (2008). Fate Tracing Reveals the Endothelial Origin of Hematopoietic Stem Cells. *Cell Stem Cell*, *3*(6), 625–636. https://doi.org/10.1016/j.stem.2008.09.018

Zuber, J., Rappaport, A. R., Luo, W., Wang, E., Chen, C., Vaseva, A. V., Shi, J., Weissmueller, S., Fellman, C., Taylor, M. J., Weissenboeck, M., Graeber, T. G., Kogan, S. C., Vakoc, C. R., & Lowe, S. W. (2011a). An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance. *Genes & Development*, *25*(15), 1628–1640. https://doi.org/10.1101/gad.17269211

Zuber, J., Shi, J., Wang, E., Rappaport, A. R., Herrmann, H., Sison, E. A., Magoon, D., Qi, J., Blatt, K., Wunderlich, M., Taylor, M. J., Johns, C., Chicas, A., Mulloy, J. C., Kogan, S. C., Brown, P., Valent, P., Bradner, J. E., Lowe, S. W., & Vakoc, C. R. (2011b). RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature*, *478*(7370), 524–528. https://doi.org/10.1038/nature10334